Izabella Thomas, Marie-Laure Betbeder

Guide de l'utilisateur de la Station Sensunique

Contenu

1 Introduction ... 3
1.1 Principes de fonctionnement ... 3
1.2 Organisation du manuel d'utilisation de la Station Sensunique ... 5
1.2.1 Conventions ... 6
2 Pré-traitements : préparation de corpus ... 6
2.1 Mise au format TEI P5 ... 6
2.2 Nettoyage manuel des fichiers ... 8
3 Connexion à la Station ... 9
4 Interface de projet ... 10
4.1 Gérer les Utilisateurs d'un projet (Onglet Utilisateurs) ... 10
4.1.1 Gestion des droits d'utilisateurs ... 10
4.1.2 Actions de l'Administrateur d'un Groupe de travail ... 13
4.2 Définir et paramétrer les projets (Onglet Projets) ... 14
4.2.1 Créer/sélectionner/ supprimer les projets (fenêtre Liste des projets) ... 15
4.3 Choisir des outils et ressources (Onglet *Outils et ressources pour la proposition d'UL*) ... 15
4.3.1 Choisir les outils (Fenêtre *Outils utilisés*) ... 16
4.3.2 Paramétrer les pondérations (Fenêtre *Pondérations*) ... 16
4.3.3 Choisir les ressources externes (Fenêtre *Ressources externes*) ... 19
4.3.4 Définir les ressources internes (Fenêtre *Ressources internes*) ... 20
4.4 Choisir les outils pour le concordancier (Onglet Outils pour le concordancier) ... 21
4.5 Définir et intégrer les corpus (Onglet *Corpus*) ... 22
4.6 Lancer l'analyse (Onglet *Analyse*) ... 24
4.7 Accéder à l'espace de travail d'un projet en cours (Onglet Analyse) ... 25
5 Analyse automatique ... 25
6 Interface de travail ... 27
6.1 Espace 1 : Visualisation de la liste d'UL/ SL ... 29
6.1.1 Onglet Liste des UL ... 29
6.1.2 Onglet Liste des SL ... 38
6.1.3 Onglet Résultats de la recherche ... 40
6.1.4 Onglet Statistiques de l'analyse ... 40
6.2 Espace 2 : Filtres sur la liste des UL/SL ... 41

6.2.1 Filtres : Valeurs requises ... 41
6.3 Espace 3 : Visualisation de fiches lexicales d'UL/SL ... 42
6.3.1 Contenu de la fiche lexicale d'une UL : modification et/ ou enrichissement manuel ... 43
6.3.2 Contenu de la fiche lexicale d'une SL : modification et / ou enrichissement manuel ... 45
6.3.3 Processus de validation des UL/SL ... 47
6.4 Espace 4 : Visualisation en contexte ... 48
6.4.1 Onglet Visualisation en corpus ... 48
6.4.2 Onglet Visualisation par phrase ... 49
6.5 Concordancier évolué ... 49
6.5.1 Expressions Régulières Sensunique ... 51
7 Export d'un dictionnaire ... 55
8 Quelques problèmes rencontrés lors de l'utilisation de la Station ... 55
9 Annexes ... 56
9.1 Liste des abréviations ... 56
9.2 Liste des catégories morphosyntaxiques Sensunique ... 57
9.3 Listes des valeurs grammaticales ... 58
9.4 Exemple d'une ressource interne au format Sensunique ... 59
Table des matières des figures ... 60
Tables des matières des tableaux ... 61

1 Introduction

Ce document décrit les fonctionnalités et l'utilisation de la **Station Sensunique**, plateforme web modulaire, collaborative et évolutive d'acquisition assistée de vocabulaire (terminologique et non terminologique), conçue à l'Université de Franche-Comté durant le projet **ANR Sensunique** (2010-2012).

La **Station Sensunique** permet :

- d'accélérer le processus d'établissement du lexique d'un domaine ou d'une Langue Contrôlée, en s'appuyant sur :
  - la collaboration de plusieurs outils TAL ;
  - l'interrogation automatique des ressources terminologiques existantes ;
  - l'intégration et l'interrogation des ressources terminologiques ou lexicales propres ;
- de gérer le lexique, de créer et d'exporter des dictionnaires.

La Station prend en entrée un corpus textuel en format XML TEI P5 et produit en sortie des dictionnaires en format XML.

1.1 Principes de fonctionnement

Comme toute plateforme terminologique, la Station intègre la mise en séquencement de plusieurs outils TAL : étiquetage, lemmatisation et extraction de termes. Sa spécificité repose sur ses autres fondements méthodologiques.

Le premier est la **multi-extraction** ou coopération de plusieurs extracteurs (Acabit, YaTea, Termostat, cf. Choisir les outils). Ce procédé donne des résultats significativement meilleurs que l'utilisation d'un seul extracteur et il permet de réduire le silence et filtrer automatiquement le bruit. Plus précisément, cumuler les résultats de 3 extracteurs de termes permet de couvrir 79 % des termes (par opposition à 58% de rappel pour le meilleur extracteur). Le meilleur moyen d'aider à déterminer le statut terminologique d'une UL est de se baser sur les résultats communs aux 2 extracteurs (YaTea et Termostat dans l'étude) avec une précision de 37 % par opposition à 28% d'un seul extracteur (Plaisantin Alecu et al. 2012).

La seconde spécificité de la Station est **le recoupement des résultats d'extraction avec des ressources lexicales** (cf. Définir les ressources internes) et **terminologiques** (IATE, TermSciences, cf. Choisir les ressources externes) existantes interrogées automatiquement. Ceci permet, d'une part, d'augmenter le potentiel terminologique d'une UL déjà recensée comme terme dans une ressource externe, et d'autre part d'attribuer un statut non-terminologique à des UL présentes dans les ressources lexicales intégrées à la Station.

Le dernier fondement méthodologique est le **calcul de trois pondérations** en fonction de diverses informations recueillies automatiquement par la Station : (1) le Poids Terminologique (PT) ou potentiel d'une UL à être un terme ; (2) le Poids de Structure Lexicale (PSL) ou potentiel d'une UL à être transformée en une Structure Lexicale ; et (3) le Poids d'Unité Lexicale (PUL) ou potentiel d'une UL à être une unité lexicale bien formée (cf. Paramétrer les pondérations). Le calcul de ces pondérations organise le travail de validation et facilite la prise de décision et l'établissement de consensus entre plusieurs analystes ou entre l'analyste et l'expert métier.

La station Sensunique fonctionne de façon modulaire, chaque module proposant à l'utilisateur plusieurs services (cf. Figure 1-1). Les modules sont organisés pour correspondre au processus d'acquisition de ressources, divisé en plusieurs étapes :

Etape 0 : Création d'un projet et gestion des utilisateurs ;

Etape 1 : Analyse automatique, qui extrait, à partir d'un corpus textuel, une liste composée d'unités terminologiques et non-terminologiques classées en fonction de leur statut et de leur potentiel terminologique ;

Etape 2 : Analyse manuelle approfondie, qui consiste en un premier filtrage de la liste opéré par l'analyste pour ne retenir que les unités potentiellement valables et un second filtrage réalisé avec l'aide de l'expert métier aboutissant à des ressources validées ;

Etape 3 : Définition des paramètres d'export et export des ressources établies.

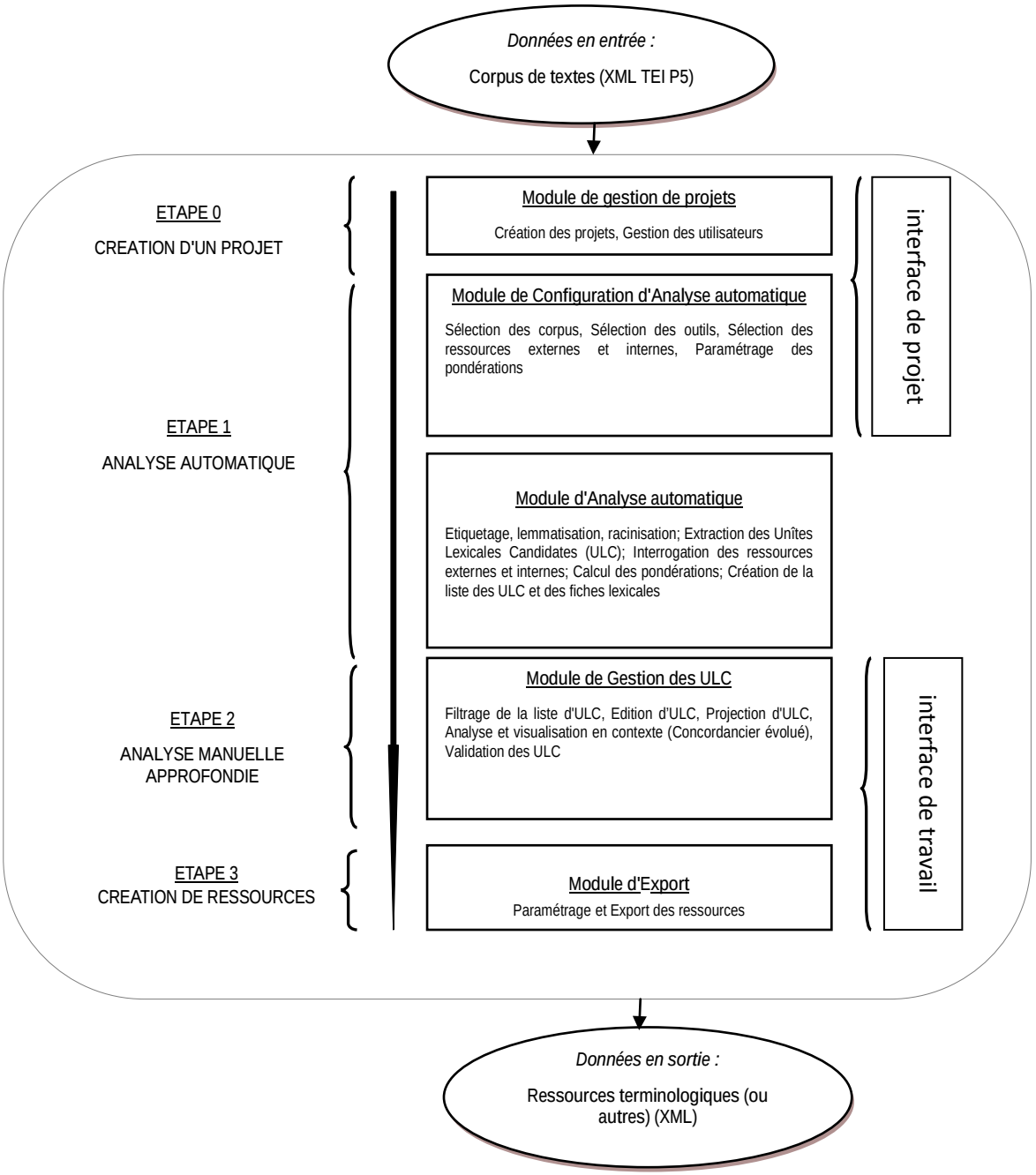

Données en entrée :
Corpus de textes (XML TEI P5)
ETAPE 0
CREATION D'UN PROJET
Module de gestion de projets
Création des projets, Gestion des utilisateurs
Module de Configuration d'Analyse automatique
Sélection des corpus, Sélection des outils, Sélection des ressources externes et internes, Paramétrage des pondérations
interface de projet
ETAPE 1
ANALYSE AUTOMATIQUE
Module d'Analyse automatique
Etiquetage, lemmatisation, racinisation; Extraction des Unîtes Lexicales Candidates (ULC); Interrogation des ressources externes et internes; Calcul des pondérations; Création de la liste des ULC et des fiches lexicales
ETAPE 2
ANALYSE MANUELLE APPROFONDIE
Module de Gestion des ULC
Filtrage de la liste d'ULC, Edition d'ULC, Projection d'ULC, Analyse et visualisation en contexte (Concordancier évolué), Validation des ULC
interface de travail
ETAPE 3
CREATION DE RESSOURCES
Module d'Export
Paramétrage et Export des ressources
Données en sortie :
Ressources terminologiques (ou autres) (XML)

Figure 1-1. Architecture de la Station Sensunique

Pour plus d'information sur la Station Sensunique :

THOMAS I., PLAISANTIN ALECU B., GERMAIN B., BETBEDER M.-L. (2014), *Station Sensunique : Architecture générale d'une plateforme web paramétrable, modulaire et évolutive d'acquisition assistée de ressources*, in Proceedings of the XVI EURALEX International Congress (Bolzano/Bozen, 15-19 July 2014) (à paraître).

THOMAS I., PLAISANTIN ALECU B., GERMAIN B., BETBEDER M.-L. (2014), *La Station Sensunique, une plateforme Web modulaire, collaborative et évolutive d'acquisition assistée de ressources terminologiques et non terminologiques (orientée Langues Contrôlées)*, in Proceedings of the XVI EURALEX International Congress (Bolzano/Bozen, 15-19 July 2014) (à paraître).

PLAISANTIN ALECU B., THOMAS I., RENAHY J. (2012), *La « multi-extraction » comme stratégie d'acquisition optimisée de ressources terminologiques et non terminologiques*, Actes de la conférence conjointe JEP-TALN-RECITAL 2012, volume 2 : TALN, ATALA/AFCP, pp. 511-518, http://www.aclweb.org/anthology/F/F12/F12-2047

Pour plus d'information sur les Langues Contrôlées :

RENAHY J., THOMAS I., CHIPPEAUX G., GERMAIN B., PETIAUX X., RATH B., DE GRIVEL V., CARDEY S., VUITTON DA., (2011), *La « langue contrôlée » et l'informatisation de son utilisation au service de la qualité des textes médicaux et de la sécurité dans le domaine de la santé*, in P. Staccini, A. Harmel, S. Darmoni, R. Gouider, Systèmes d'information pour l'amélioration de la qualité en santé. Comptes rendus des quatorzièmes Journées francophones d'informatique médicale, Tunis, 23 et 24 septembre 2011 (coll. Informatique et Santé), Springer-Verlag, http://tesniere.univ-fcomte.fr/ressources/Renahy_JFIM2011.pdf.

RENAHY J., DEVITRE D., THOMAS I., DZIADKIEWICZ A., (2009), *Controlled language norms for the redaction of security protocols: finding the median between system needs and user acceptability*, in Proceedings of the 11th International Symposium on Social Communication, Santiago de Cuba, Cuba, 19-23 January 2009, pp. 289-293.

RENAHY J., THOMAS I., (2009), *Compagnon LiSe : A Collaborative Controlled Language Writing Assistant*, in ISMTCL Proceedings, International Review BULAG, PUFC, ISSN 0758 6787, ISBN 978-2-84867-261-8, pp. 223-230.

1.2 Organisation du manuel d'utilisation de la Station Sensunique

Ce manuel a pour objectif de décrire les différents services proposés par la Station Sensunique. Il est organisé suivant le processus chronologique de l'acquisition des ressources et selon les différentes interfaces de la Station. Il décrit les actions successives de l'utilisateur (et du système) :

- préparation du corpus (Chapitre 3, Prétraitement : préparation de corpus) ;
- connexion à la Station (Chapitre 4, Connexion à la Station) ;
- création de projets et configuration de l'analyse automatique à partir de l'Interface de projet (Chapitre 5, Interface de projet) ;
- résultats de l'analyse automatique (Chapitre 6, Analyse automatique) ;
- analyse manuelle approfondie et export des données à partir de l'Interface de travail (Chapitre 7, Interface de travail).
- export des ressources (Chapitre 8, Export d'un dictionnaire).

Note

La Station Sensunique est un outil issu d'un projet de recherche. Il n'est en aucun cas un outil commercial. De ce fait, nous ne pouvons pas garantir son fonctionnement parfait (voir la liste de problèmes rencontrés).

Toute demande ou remarque sur le fonctionnement de la Station Sensunique ou sur d'éventuels problèmes concernant son utilisation peuvent être adressées à *izabella.thomas@univ-fcomte.fr*.

1.2.1 Conventions

Les opérations qui peuvent être effectuées par l'utilisateur sont décrites de la façon suivante :

Nom de l'opération à effectuer *Actions utilisateurs*

> action à effectuer > action à effectuer … ;

➔ résultat intermédiaire ;

> action à effectuer > action à effectuer …;

résultat final.

Exemple :

Créer un nouveau projet (fenêtre *Liste de projet*) *Actions utilisateurs*

> cliquer sur *Créer un nouveau projet* > remplir la fenêtre de dialogue (donner un nom au projet) > cliquer sur *Ajouter le projet* ;

➔ le projet apparait dans la liste des projets avec la date de création ;

> sélectionner le projet pour continuer .

L'icône introduit des informations complémentaires par rapport à des opérations à effectuer.

2 Pré-traitements : préparation de corpus

Les corpus doivent être préparés en dehors de la Station. Ils peuvent contenir un ou plusieurs fichiers, converti(s) au format XML TEI P5 et nettoyé(s) manuellement. L'intégration des corpus à la Station Sensunique est décrite dans la section Définir et intégrer les corpus (Onglet Corpus).

2.1 Mise au format TEI P5

Il existe plusieurs logiciels de conversion de format de fichiers. La méthode la plus simple consiste à utiliser OxGarage (http://oxgarage.oucs.ox.ac.uk:8080/ege-webclient/), un service en ligne permettant de convertir une grande variété de formats de documents, qu'il s'agisse de textes, de présentations ou de feuilles de calcul.

La liste des formats pris en charge à ce jour par OxGarage :

Documents textes :

- Compiled TEI ODD Document
- DocBook Document
- Microsoft Word (.doc)

- Microsoft Word (.docx)
- ODD Document
- Open Office Text (.odt)
- OpenOffice 1.0 Text (.sxw)
- OpenOffice Text (.odt)
- Plain Text (.txt)
- Rich Text Format (.rtf)
- TCP XML Document
- TEI P4 XML Document
- TEI P5 XML Document
- TEI Tite XML Document
- WordPerfect (.wpd)
- xHTML

Présentations :

- Microsoft PowerPoint (.ppt)
- Microsoft PowerPoint (.pptx)
- Open Office (.odp)
- OpenOffice.org 1.0 (.sxi)

Feuilles de calcul :

- Comma-Separated Values (.csv)
- Microsoft Excel (.xls)
- Microsoft Excel (.xlsx)
- Open Office (.ods)
- OpenOffice.org 1.0 (.sxc)
- Tab-Separated Values (.tsv)

La procédure de conversion consiste à choisir le format de fichier source (à convertir) et le format du fichier cible (après conversion), à télécharger le fichier source et à récupérer le fichier cible converti en format XML TEI P5 (cf. Figure 2-1). Ce processus est très simple, par contre, il arrive que le fichier converti contienne des caractères corrompus, qui ne peuvent pas être correctement gérés par la Station Sensunique. Dans ce cas, un nettoyage manuel des fichiers convertis est vivement conseillé.

Pour plus d'information sur le format TEI Lite :

DTD TEI Lite :
http://www.tei-c.org/release/xml/tei/custom/schema/dtd/tei_lite.dtd

Documentation TEI Lite (anglais) :
http://www.tei-c.org/Vault/P5/2.1.0/doc/tei-p5-exemplars/html/teilite.doc.html

Documentation TEI Lite (français) :
http://www.tei-c.org/Guidelines/Customization/Lite/teiu5_fr.html

TEI OpenOffice Package :
http://wiki.tei-c.org/index.php/TEI_OpenOffice_Package

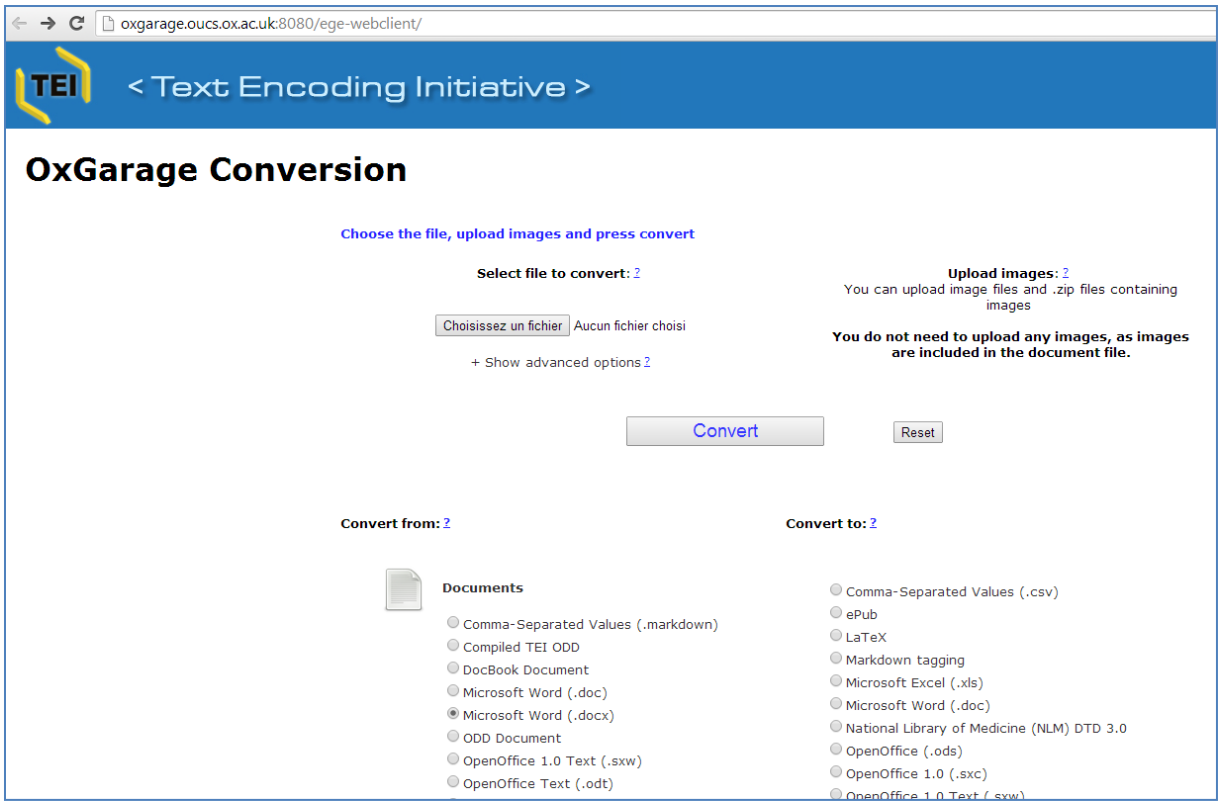


Figure 2-1. Oxgarage

2.2 Nettoyage manuel des fichiers

Il est conseillé de réviser et nettoyer manuellement les fichiers convertis en format TEI P5. Il arrive que certains caractères soient corrompus lors de la conversion. Pour être en mesure de visualiser les caractères corrompus et de les effacer, il faut afficher les fichiers convertis dans un éditeur de texte avancé, tel que Notepad++. Les caractères à effacer sont alors facilement identifiables, car ils apparaissent sous forme de carrés ou de points (un peu comme des puces dans Word, voir Figure 2 - 2). Bien que cette méthode soit fastidieuse, et que le risque d'omission soit important, nous ne disposons pas pour le moment de meilleur procédé pour assurer l'intégrité des fichiers convertis.

```
   <hi rend="bold">Description du problème Causes Solutions</hi
</p>
<p>a. Paramètres du FACS mal ajustés. a. Vérifier les paramètre
<p>sang frais.</p>
<p>1. Dot plot anormal</p>
<p>b. Le sang total n'est pas complètement</p>
<p>lysé. (Voir les figures 5a, 5b et 5c.)</p>
<p>b. Vérifier le tampon et les conditions de lyse. S'ils</p>
<p>sont corrects :</p>
<p>□□centrifuger l'échantillon et répéter la lyse.</p>
<p>□□Augmenter la quantité de tampon de lyse à 4</p>
<p>ml.</p>
<p>□□Augmenter la température d'incubation de lyse</p>
<p>à 37 □C.</p>
<p>□□Augmenter le temps d'incubation de lyse à</p>
<p>20 minutes.</p>
<p>(Remarque : Pour optimiser la lyse, les étapes cidessus</p>
<p>peuvent être effectuées séparément ou en</p>
<p>même temps.)</p>
<p>a. Le cytomètre de flux doit être aligné. a. Vérifier les co
<p>s'assurer qu'ils sont dans la plage correcte.</p>
```

Figure 2-2. Exemple d'un fichier converti en format TEI P5 contenant des caractères corrompus

La Station Sensunique ne gère pas correctement les fichiers contenant les caractères corrompus. La plupart des échecs d'analyse analyses sont dus à la présence de caractères corrompus dans le corpus.

3 Connexion à la Station

Pour se connecter à la Station Sensunique, il faut disposer d'identifiants (login et mot de passe), qui sont délivrés par l'Administrateur de la Station lors de la première inscription (cf. Figure 3-1).

Pour obtenir les identifiants, il faut :

- remplir et envoyer la Demande d'inscription (par mail) ;
- prendre connaissance et signer la Charte des Utilisateurs de la Station Sensunique. La Charte des Utilisateurs, signée, doit être envoyée à :

UNIVERSITÉ DE FRANCHE-COMTÉ
Direction de la Valorisation
Temis Innovation - Maison des Microtechniques
18, rue Alain Savary
25000 BESANÇON

Le droit d'utilisation est accordé à un Administrateur d'un Groupe de Projet (cf. Gestion des droits d'Utilisateurs) pour une durée de un (1) an à compter de la création du compte utilisateur de l'Administrateur, renouvelable par tacite reconduction d'année en année. Par la suite, l'Administrateur d'un Groupe de Projet a le droit de créer des comptes à d'autres utilisateurs de la Station, comptes rattachés à son Groupe de Projet (cf. Gestion des droits d'Utilisateurs).

Connexion à la station
Identifiant* :
Mot de passe* :
Rester connecté pendant 1 semaine
Connexion
Demande d'inscription
Nom* :
Prénom* :
Organisme* :
Laboratoire* :
Email* :
Téléphone :
Utilisation* : Enseignement Recherche
Je confirme avoir pris connaissance de la Charte des utilisateurs.
Envoyer la demande
Après renvoi de la convention signée, vous recevrez un message avec vos nom d'utilisateur et mot de passe.
Avertissement
La station Sensunique n'est utilisable qu'après insciption. L'utilisation est gratuite pour des fins de recherche ou d'enseignement.
Pour toute question, contactez Izabella Thomas.
Informations
La Station Sensunique est une plateforme web modulaire, collaborative et évolutive d'acquisition assistée de vocabulaire (terminologique et non terminologique). Elle a été conçue à l'Université de Franche-Comté durant le projet ANR Sensunique (2010-2012).
La Station Sensunique permet :
d'accélérer le processus d'établissement du lexique d'un domaine ou d'une Langue Contrôlée, en s'appuyant sur :
la collaboration de plusieurs outils TAL ;
l'interrogation automatique des ressources terminologiques existantes ;
l'intégration et l'interrogation des ressources terminologiques ou lexicales propres ;
de gérer le lexique, de créer et d'exporter des dictionnaires.
Télécharger le Guide de l'utilisateur de la Station Sensunique
Outils et ressources intégrés
Extraction de termes :
Acabit
TermoStat
YaTeA
Racinisation :
Lingua::Stem
Etiquetage morphosyntaxique :
Brill
TreeTagger
Ressources intégrées :
TermSciences (INIST)
IATE (Base de données terminologique de l'Union Européenne)
Lemmatisation :
TreeTagger
Flemm V2
Flemm V3.1

Figure 3-1. Interface de connexion à la Station Sensunique

4 Interface de projet

L'interface de projet est composée de deux onglets :

- Onglet Utilisateurs, qui permet de gérer les utilisateurs d'un groupe de travail;
- Onglets Projets, qui permet de définir les projets, paramétrer et lancer l'analyse automatique sur les projets sélectionnés.

Les actions liées à la gestion de projet ne peuvent être effectuées que par l'Administrateur du Groupe et/ou Responsable de projet (cf. Gestion des droits d'Utilisateurs).

4.1 Gérer les Utilisateurs d'un projet (Onglet Utilisateurs)

4.1.1 Gestion des droits d'utilisateurs

La Station Sensunique est organisée de manière hiérarchique (cf. Figure 4-1). Elle est gérée par un Administrateur de la Station qui crée des Groupes de travail et nomme un Administrateur pour chaque Groupe de travail. L'Administrateur du Groupe de travail crée des Projets et nomme un Responsable de chaque Projet.

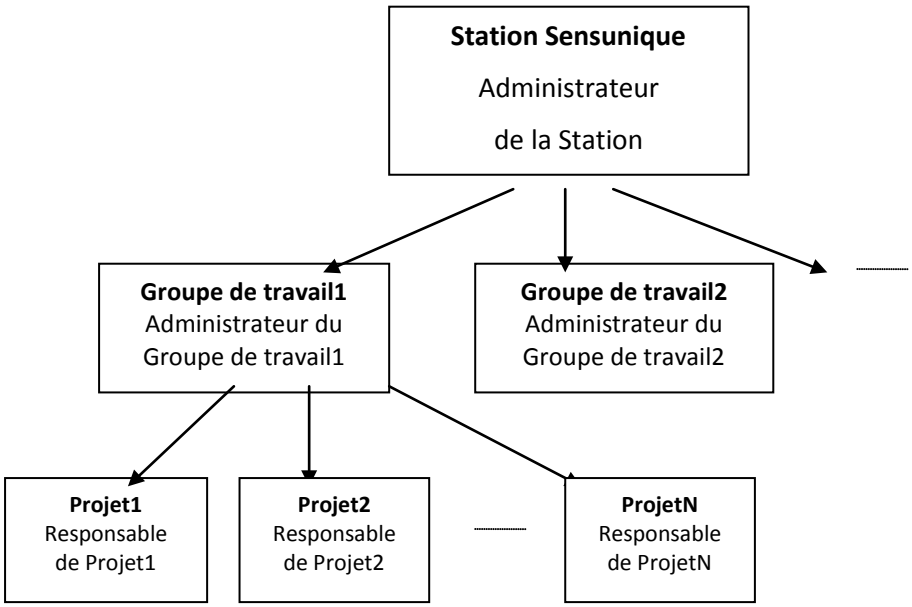


Figure 4-1. Gestion des utilisateurs

Groupe de travail

Le Groupe de travail est créé suite à la réception de la Charte des Utilisateurs signée par l'Administrateur du Groupe de travail. Il est composé d'un ou de plusieurs Utilisateur(s) appartenant à une structure signataire de la Charte des Utilisateurs. Le Groupe de travail est géré par l'Administrateur du groupe de travail.

Utilisateur

Un Utilisateur est une personne rattachée à un Groupe de travail. L'Utilisateur peut avoir le statut d'Administrateur du groupe de travail, de Responsable du projet, ou de Participant (cf. Tableau 4.1).

Administrateur du Groupe de travail

L'Administrateur du Groupe de travail est nommé par l'Administrateur de la Station ; il n'existe qu'un seul Administrateur par Groupe de travail. Il gère les Utilisateurs et les Projets du Groupe de travail. Il détient plusieurs droits spécifiques :

- il crée et gère des comptes Utilisateurs pour chaque personne qu'il souhaite intégrer à son Groupe de travail ;
- il crée des Projets pour son Groupe de travail ;
- il affecte des Utilisateurs à un Projet et leur attribue le statut de Responsable du Projet ou de Participant ;
- il peut créer /sélectionner/ supprimer des Corpus de chaque Projet de son Groupe de travail ;
- il peut lancer/supprimer l'analyse de chaque Projet de son Groupe de travail.

L'Administrateur du groupe signe la Charte des Utilisateurs et est légalement responsable de tous les Utilisateurs de son groupe.

Responsable du Projet

Le Responsable du Projet est nommé par l'Administrateur du Groupe ; il n'existe qu'un Responsable du Projet par projet. Il a plusieurs droits spécifiques :

- il affecte des Participants au Projet dont il est responsable (parmi les Utilisateurs déclarés dans le Groupe de travail) ;
- il peut créer /sélectionner/ supprimer des Corpus du Projet dont il est responsable ;

- il peut lancer/effacer l'analyse du Projet dont il est responsable.

Le Responsable du Projet **ne peut pas** :

- ajouter des nouveaux Utilisateurs à son Groupe de travail ;
- créer de nouveaux Projets pour son Groupe de travail.

Participant

Le statut de Participant est affecté soit par l'Administrateur du groupe, soit par le Responsable du Projet. Le Participant possède les droits suivants :

- il a les droits de lecture et d'écriture sur tous les Projets du Groupe auquel il appartient.

Le Participant **ne peut pas** :

- créer de nouveaux Projets pour son Groupe de travail ;
- ajouter des nouveaux Utilisateurs à son Groupe de travail ;
- affecter des Utilisateurs à un Projet ;
- créer /sélectionner/ supprimer de Corpus du projet auquel il est affecté ;
- lancer/effacer l'analyse de son Projet.

Projet

Un Groupe de travail peut contenir un ou plusieurs Projet(s). Un Projet contient au moins un Corpus d'Analyse (cf. Définir et intégrer les corpus).

Droits	Périmètre d'action	Administrateur Station	Administrateur Groupe de travail	Responsable Projet	Participant
Créer/ supprimer un Groupe de travail	Station Sensunique	o	n	n	n
Nommer l'Administrateur du Groupe de travail	Groupe de travail	o	n	n	n
Créer/ supprimer et gérer les comptes d'Utilisateurs	Groupe de travail	n	o	n	n
Créer/supprimer des Projets	Groupe de travail	n	o	n	n
Nommer le Responsable du Projet	Projet	n	o	n	n
Créer /sélectionner/ supprimer un Corpus d'un Projet de son Groupe de travail	Groupe de travail	n	o	n	n
Lancer/effacer l'analyse de chaque Projet de son Groupe de travail	Groupe de travail	n	o	n	n
Affecter les Utilisateurs à un Projet (parmi les utilisateurs d'un Groupe de travail)	Projet	n	o	o	n
Créer /sélectionner/ supprimer de Corpus du Projet auquel il est affecté	Projet	n	o	o	n
Lancer/effacer l'analyse du Projet auquel il est affecté	Projet	n	o	o	n
Lire et écrire sur tous les Projets du Groupe de travail auquel il appartient	Groupe de travail	n	o	o	o

Légende : o =oui, n =non

Tableau 4-1. Récapitulatif des droits selon le type d'utilisateur

4.1.2 Actions de l'Administrateur d'un Groupe de travail

L'Administrateur d'un Groupe de travail gère les Utilisateurs à partir de l'onglet *Utilisateurs*, dans l'Interface de projet (cf. Figure 4-2).

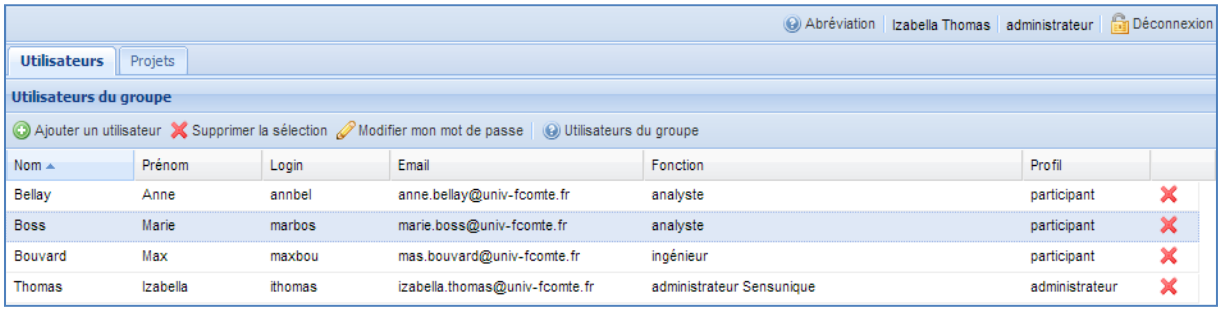


Figure 4-2. Interface de projet, Onglet *Utilisateurs*

Si vous êtes Administrateur d'un Groupe de travail, vous pouvez :

Ajouter un nouvel utilisateur à votre Groupe *Actions utilisateurs*

> cliquer sur *Ajouter un utilisateur* > remplir la fenêtre de dialogue > confirmer ;

➔ le nouvel utilisateur apparait dans la liste.

Supprimer une sélection d'utilisateurs de votre Groupe

> appuyer sur la touche CTRL pour sélectionner plusieurs utilisateurs à supprimer > cliquer sur *Supprimer la sélection* > confirmer ;

➔ les utilisateurs sélectionnés disparaissent de la liste.

Modifier le mot de passe de chaque Participant de votre Groupe

> sélectionner le Participant > remplir la fenêtre de dialogue > confirmer.

Si vous êtes Participant, vous pouvez modifier votre mot de passe.

Le champ Fonction correspond à la véritable fonction d'un utilisateur dans le groupe de travail ; il est rempli librement. Le champ Profil correspond à un statut prédéfini dans la station (cf. Gestion des droits d'Utilisateurs).

En double-cliquant sur le nom d'un champ, on effectue un tri alphabétique croissant ou décroissant.

4.2 Définir et paramétrer les projets (Onglet Projets)

L'onglet *Projets* permet de définir les projets et de paramétrer l'analyse automatique (cf. Figure 4-3).

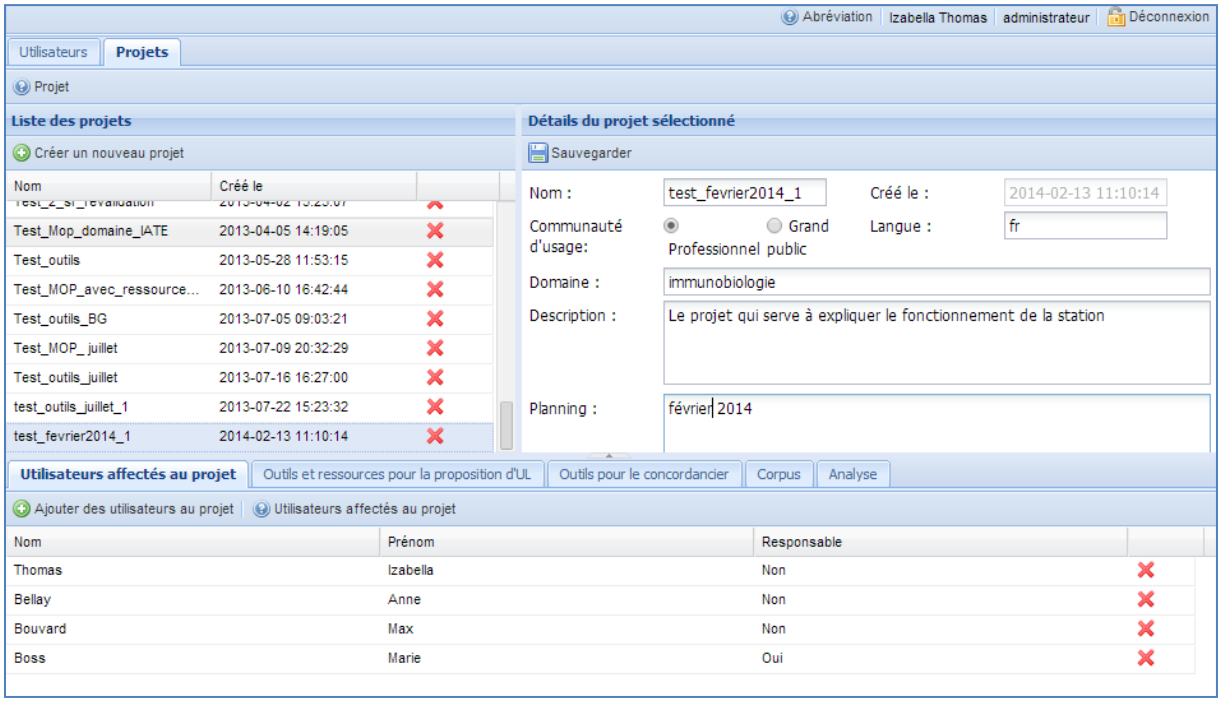


Figure 4-3. Onglet Projets, vue générale

4.2.1 Créer/sélectionner/ supprimer les projets (fenêtre Liste des projets)

Si vous êtes Administrateur du Groupe de travail, vous pouvez :

Créer un nouveau projet (fenêtre Liste de projet) *Actions utilisateurs*

> cliquer sur *Créer un nouveau projet* > remplir la fenêtre de dialogue (donner un nom au projet) > cliquer sur *Ajouter le projet* ;

➔ le projet apparait dans la liste des projets avec la date de création ;

> sélectionner le projet pour continuer .

Décrire le nouveau projet (fenêtre *Détails du projet sélectionné*)

> remplir les champs demandés > cliquer sur *Sauvegarder*.

Il faut remplir tous les champs demandés pour pouvoir sauvegarder les informations.

Nommer un Responsable du projet (Onglet *Utilisateurs affectés au projet*)

> sélectionner un utilisateur > dans la colonne *Responsable* double-cliquer sur oui/non > choisir la valeur oui/non.

Il ne peut y avoir qu'un seul Responsable par projet; le statut de Responsable ouvre des droits particulier à l'utilisateur sélectionné comme Responsable du projet (cf. Gestion des droits d'Utilisateurs).

Si vous êtes Administrateur du Groupe de travail ou Responsable du projet, vous pouvez :

Affecter les utilisateurs au projet (Onglet *Utilisateurs affectés au projet*) *Actions utilisateurs*

> cliquer sur *Ajouter des utilisateurs au projet* > choisir des utilisateurs dans la liste > cliquer sur *Ajouter les utilisateurs sélectionnés au projet* ;

➔ les utilisateurs sélectionnés apparaissent dans la liste des utilisateurs.

Retirer des utilisateurs d'un projet (Onglet *Utilisateurs affectés au projet*)

> cliquer sur l'icône dans la ligne de l'utilisateur à retirer > confirmer ;

➔ l'utilisateur sélectionné disparait de la liste.

4.3 Choisir des outils et ressources (Onglet *Outils et ressources pour la proposition d'UL*)

L'onglet *Outils et ressources pour la proposition d'UL* (cf. Figure 4-4) permet de configurer l'analyse automatique, c'est-à-dire choisir les outils et les ressources (externes et internes) à utiliser durant son exécution, et configurer les pondérations. Il est composé de 4 fenêtres :

- *Outils utilisés* (cf. Choisir les outils) ;
- *Pondérations* (cf. Paramétrer les pondérations) ;
- *Ressources externes* (cf. Choisir les ressources externes);
- *Ressources internes* (cf. Choisir les ressources internes).

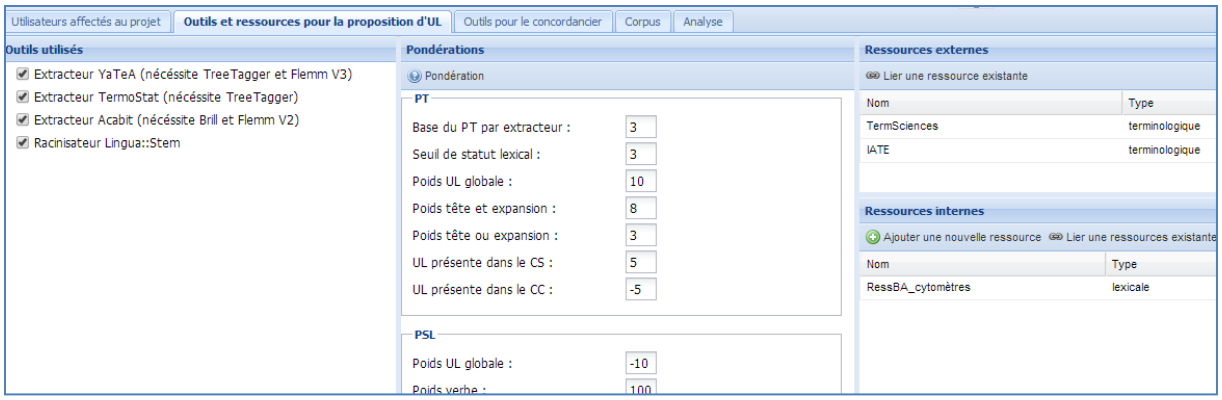


Figure 4-4. Onglet Outils et ressources pour la proposition d'UL

4.3.1 Choisir les outils (Fenêtre *Outils utilisés*)

La Station Sensunique intègre les outils suivants :

- les étiqueteurs morphosyntaxiques Treetagger (Schmid, 1994) et Brill[1] (Brill, 1992) ;
- l'analyseur flexionnel du français Flemm v2 et v3 (Namer, 2000) ;
- les extracteurs de termes Acabit (Daille, 1994), TermoStat (Drouin, 2003) et YaTeA (Aubin et al. 2006) ;
- le racinisateur Lingua:: Stem[2].

L'analyste peut sélectionner de 1 à 3 chaînes d'outils parmi : (1) TreeTagger - Termostat ; (2) Brill - Flemm v2 - Acabit ; (3) TreeTagger - Flemm v3 - YaTeA. Bien que la sélection d'une seule chaîne suffise, la Station est optimisée avec l'emploi des 3 chaînes grâce au procédé de multi-extraction (Plaisantin Alecu et al., 2012). Le racinisateur Lingua::Stem permet de récupérer les relations dérivationnelles entre les Unités Lexicales (UL).

Si vous êtes Administrateur du Groupe de travail ou Responsable du projet, vous pouvez :

Sélectionner les outils *Actions utilisateurs*

> cocher les outils dans la fenêtre *Outils utilisés* > cliquer sur *Enregistrer cette configuration*.

4.3.2 Paramétrer les pondérations (Fenêtre *Pondérations*)

Pour chaque UL, la Station calcule 3 types de poids : Poids de potentiel Terminologique (PT), Poids de Structure Lexicale (PSL), Poids d'Unité Lexicale (PUL). Ces poids sont calculés en fonction des critères énumérés dans les tableaux-ci-dessus.

La Station propose des valeurs de pondération par défaut ; elles peuvent néanmoins être ajustées par l'analyste en fonction de chaque projet (cf. Figure 4-5).

[1] Avec le lexique et le fichier de règles fournis par l'ATILF-CNRS, de Nancy.
[2] http://search.cpan.org/~sdp/Lingua-Stem-Fr0.02/lib/Lingua/Stem/Fr.pm

PT (Poids de potentiel Terminologique) : détermine si une UL peut être considérée comme un terme.

Paramètres de pondération	Explication
Nombre d'extracteurs ayant proposé une UL	
Base du PT par extracteur	Poids par extracteur ayant proposé une UL ; au carré pour 2 extracteurs, au cube pour 3 extracteurs. ***Exemple :*** *Pour la base du PT = 3, une UL attestée par 1 extracteur aura le poids de 3, par deux extracteurs 3^2= 9, par 3 extracteurs 3^3= 27*
Seuil de statut terminologique	Seuil à partir duquel une UL est considérée comme un terme
Attestation par une ressource terminologique	
Poids UL globale	Poids attribué à une UL lorsqu'elle est attestée par une ressource terminologique externe
Poids tête et expansion	Poids attribué à une UL lorsque sa Tête **et** son Expansion sont attestées par une ressource terminologique externe
Poids tête ou expansion	Poids attribué à une UL lorsque sa Tête **ou** son Expansion sont attestées par une ressource terminologique externe
Attestation dans un autre corpus	
UL présente dans le CS	Poids attribué à une UL présente dans le Corpus Support
UL présente dans le CC	Poids (négatif) attribué à une UL présente dans le Corpus Contrastif

Tableau 4-2. Paramètres de pondération du PT

PSL (Poids de Structure Lexicale) : détermine si une UL peut être considérée comme une Structure Lexicale.

Paramètres de pondération	Explication
Attestation dans une ressource terminologique	
Poids UL globale	Poids (négatif) attribué à une UL lorsqu'elle est attestée dans une ressource terminologique externe
Matrice morphosyntaxique	
Poids verbe	Poids attribué à une UL dont la matrice morphosyntaxique est ou contient un verbe (Ver)
Poids participe	Poids attribué à une UL dont la matrice morphosyntaxique est ou contient un participe passé (ou présent) adjectival (Vppe ou Vppr)
Densité de la famille dérivationnelle	
Seuil d'UL dérivées	Seuil à partir duquel le poids d'UL dérivées est attribué
Poids d'UL dérivées	Poids attribué si le nombre d'UL dérivées distinctes de l'UL analysée dépasse le seuil *Exemple :* *A partir de 3 UL dérivées (seuil), on attribue le poids de 6 à l'UL analysée*
Densité de collocations	
Seuil d'UL collocatives	Seuil à partir duquel le poids d'UL collocatives est attribué
Poids d'UL collocatives	Poids attribué si le nombre d'UL collocatives distinctes de l'UL analysée dépasse le seuil *Exemple :* *A partir de 3 UL collocatives (seuil), on attribue le poids de 6 à l'UL analysée*
Extraction par Acabit	
Poids Acabit	Poids attribué à une UL extraite par Acabit

Tableau 4-3. Paramètres de pondérations du PSL

PUL (Poids d'Unité Lexicale) : détermine si une UL peut être considérée comme une UL.

Paramètres de pondération	Explication
Nombre d'extracteurs ayant proposé une UL	
Seuil de nombre d'extracteurs	Seuil à partir duquel le poids du nombre d'extracteurs est attribué
Poids du nombre d'extracteurs	Poids attribué à une UL lorsque le nombre d'extracteurs dépasse le seuil

Tableau 4-4. Paramètres de pondération du PUL

Si vous êtes Administrateur du Groupe de travail ou Responsable du projet, vous pouvez :

Changer les valeurs de pondérations *Actions utilisateurs*

> modifier les valeurs de pondérations (ou laisser les valeurs par défaut) > cliquer sur *Enregistrer la configuration*.

Pondérations
Pondération
PT
Base du PT par extracteur : 3
Seuil de statut lexical : 3
Poids UL globale : 10
Poids tête et expansion : 8
Poids tête ou expansion : 3
UL présente dans le CS : 5
UL présente dans le CC : -5
PSL
Poids UL globale : -10
Poids verbe : 100
Poids adjectival : 10
Seuil de dérivée: 3
Poids de dérivée : 3
Seuil collocative : 3
Poids collocative : 3
Poids Acabit: 3
PUL
Nombre d'extracteur : 2
Poids du nombre d'extracteur : 5

Figure 4-5. Valeurs de pondérations

4.3.3 Choisir les ressources externes (Fenêtre *Ressources externes*)

2 ressources externes, prédéfinies, sont intégrées à la Station Sensunique :

- TermSciences[3], portail terminologique multidisciplinaire développé par CNRS-INIST ;

- IATE[4], base de données terminologique de l'Union Européenne.

[3] http://www.termsciences.fr/
[4] http://iate.europa.eu/iatediff/SearchByQueryLoad.do?method=load

L'interrogation automatique par web service de ces deux ressources externes prédéfinies influe sur les pondérations et participe à l'enrichissement des informations rattachées à chaque ULC, telles que définitions et synonymes. L'analyste peut choisir de les intégrer ou non à l'analyse ; il peut aussi restreindre leur interrogation à un domaine ou un sous-domaine précis (cf. Figure 4-6).

Si vous êtes Administrateur du Groupe de travail ou Responsable du projet, vous pouvez :

Lier une ressource externe au projet *Actions utilisateurs*

> cliquer sur *Lier une ressource existante* > choisir la ressource dans la liste déroulante > attribuer à la ressource un **Indice de Pondération*.**

➔ la ressource choisie apparait dans la liste.

> choisir le domaine :

- **IATE** : > cliquer sur le champ *Domaine* > choisir le domaine dans la liste déroulante.
- **TermSciences** : le domaine ne peut pas être choisi pour TermSciences.

***Indice de Pondération** : le poids qui sera attribué à une UL si elle est attestée dans une ressource.

Retirer une ressource externe du projet *Actions utilisateurs*

> cliquer sur l'icône à côté de la ressource à retirer > confirmer le retrait de la ressource.

Ressources externes

Lier une ressource existante

Nom	Type	Domaine	Indice	
TermSciences	terminologique	VIE POLITIQUE	10	
IATE	terminologique	VIE POLITIQUE	15	

Figure 4-6. Ressources externes

4.3.4 Définir les ressources internes (Fenêtre *Ressources internes*)

En plus de ressources externes prédéfinies, la Station permet d'intégrer à chaque nouveau projet d'autres ressources spécifiques, moyennant leur mise au format prédéfini dans la Station (cf. Figure 4-7). Il peut s'agir aussi bien de ressources terminologiques (e.g. des dictionnaires spécialisés) qui augmentent le potentiel terminologique des ULC, que des ressources non-terminologiques, qui augmentent le poids d'unité lexicale d'une ULC tout en diminuant son potentiel terminologique (cf. Paramétrer les pondérations). Par ailleurs, des ressources constituées au préalable dans la Station, résultant d'autres projets, peuvent aussi être intégrées en tant que ressources internes.

La mise en format des ressources doit être effectuée en dehors de la Station. Le format de ressources pouvant être intégrées à la Station Sensunique est explicitée dans le fichier Excel à télécharger sur le site de la Station Sensunique. La forme définitive des ressources à intégrer (exemple en Annexe 9.4) est obtenue par enregistrement de ce fichier au format txt (*Enregistrer sous/Texte (séparateur : tabulation)*).

Si vous êtes Administrateur du Groupe de travail ou Responsable du projet, vous pouvez :

Ajouter une nouvelle ressource interne au projet *Actions utilisateurs*

> cliquer sur *Ajouter une nouvelle ressource* > remplir les champs demandés > télécharger le fichier source (au format Sensunique) > attribuer un Indice de Pondération* à la ressource > cliquer sur *Ajouter la ressource* ;

➔ la ressource apparait dans la liste des ressources.

Lier une ressource interne existante au projet *Actions utilisateurs*

> cliquer sur *Lier une ressource existante* > choisir la ressource dans la liste déroulante > attribuer un **Indice de Pondération*** à la ressource > cliquer sur *Lier la ressource* ;

➔ la ressource apparait dans la liste des ressources.

***Indice de Pondération** : le poids qui sera attribué à une UL si elle est attestée dans une ressource.

Retirer une ressource interne du projet

> cliquer sur l'icône à côté de la ressource à retirer > confirmer le retrait de la ressource.

Supprimer une ressource interne de la Station

> cliquer sur l'icône à côté de la ressource à supprimer > confirmer la suppression de la ressource.

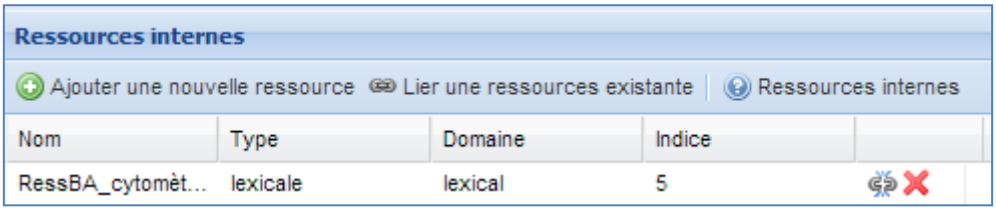


Figure 4-7. Ressources internes

4.4 Choisir les outils pour le concordancier (Onglet Outils pour le concordancier)

La Station Sensunique intègre un concordancier avancé, permettant plusieurs types de recherche sur le Corpus d'Analyse (cf. Concordancier évolué). Cet onglet permet de présélectionner les outils qui pourront être utilisés par le concordancier lors de recherches sur le corpus analysé (cf. Figure 4-8).

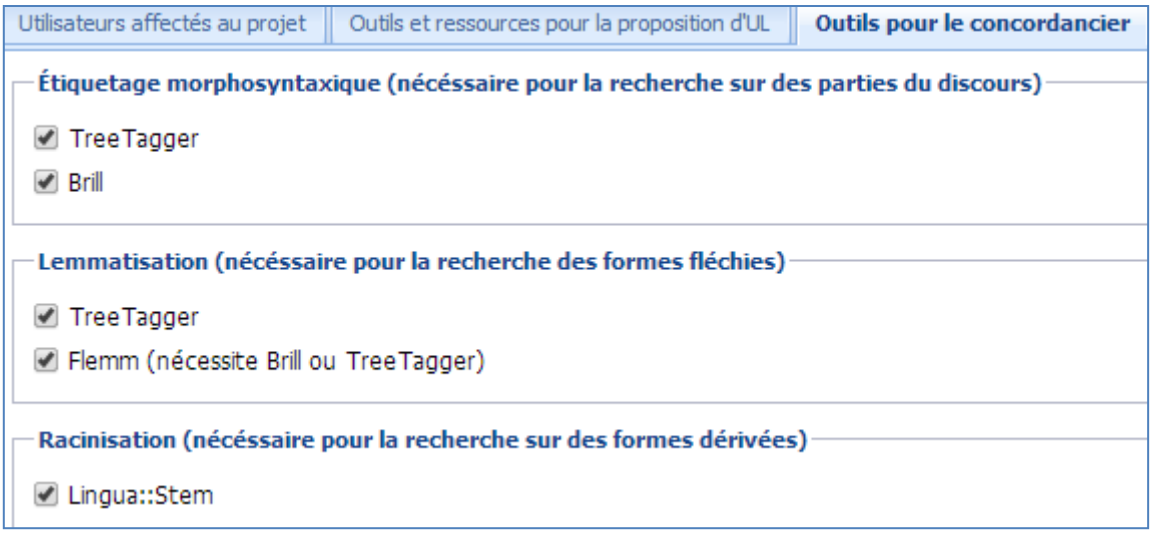

Utilisateurs affectés au projet
Outils et ressources pour la proposition d'UL
Outils pour le concordancier
Étiquetage morphosyntaxique (nécéssaire pour la recherche sur des parties du discours)
TreeTagger
Brill
Lemmatisation (nécéssaire pour la recherche des formes fléchies)
TreeTagger
Flemm (nécessite Brill ou TreeTagger)
Racinisation (nécéssaire pour la recherche sur des formes dérivées)
Lingua::Stem

Figure 4-8. Outils pour le concordancier

Si vous êtes Administrateur du Groupe de travail ou Responsable du projet, vous pouvez :

Sélectionner les outils pour le concordancier *Actions utilisateurs*

> cocher les outils dans la fenêtre *Outils pour le concordancier* > cliquer sur *Enregistrer cette configuration*.

4.5 Définir et intégrer les corpus (Onglet *Corpus*)

Pour le même Projet, trois types de corpus textuels peuvent être simultanément analysés par la Station :

- le **Corpus d'Analyse (CA)** : c'est un corpus obligatoire duquel sont extraites les UL à analyser ;
- le **Corpus Support (CS**) : c'est un corpus facultatif, du même domaine que le CA. En recoupant les UL retrouvées dans les deux corpus (CA et CS), l'algorithme de pondération renforce leur potentiel terminologique. Ce procédé est inspiré de l'hypothèse de Drouin (2003) prouvant qu'une UL extraite de deux corpus différents du même domaine a plus de probabilité d'être un terme du domaine ;
- le **Corpus Contrastif (CC**) : c'est un corpus facultatif, contenant des textes généralistes, non relatifs au domaine analysé. L'exploitation d'un CC permet à l'algorithme de pondération d'augmenter la qualité des résultats en diminuant le potentiel terminologique des UL issues du CA et du CC à la fois. Ce procédé est également inspiré de Drouin (2003) qui prouve qu'une UL extraite d'un corpus de domaine et d'un corpus généraliste a plus de probabilité d'être une unité du lexique général qu'un terme du domaine.

Les corpus sont (ré)utilisables dans plusieurs projets. En outre, un corpus n'est pas intrinsèquement lié à un statut particulier (CA, CS ou CC) : ce statut lui est attribué en fonction du projet, par un analyste. Par conséquent, le même corpus peut être utilisé comme un CA dans un projet particulier et comme un CC dans un autre projet. Ceci permet une meilleure exploitation des différents corpus constitués dans un groupe de travail ayant des projets différents (cf. Figure 4-9).

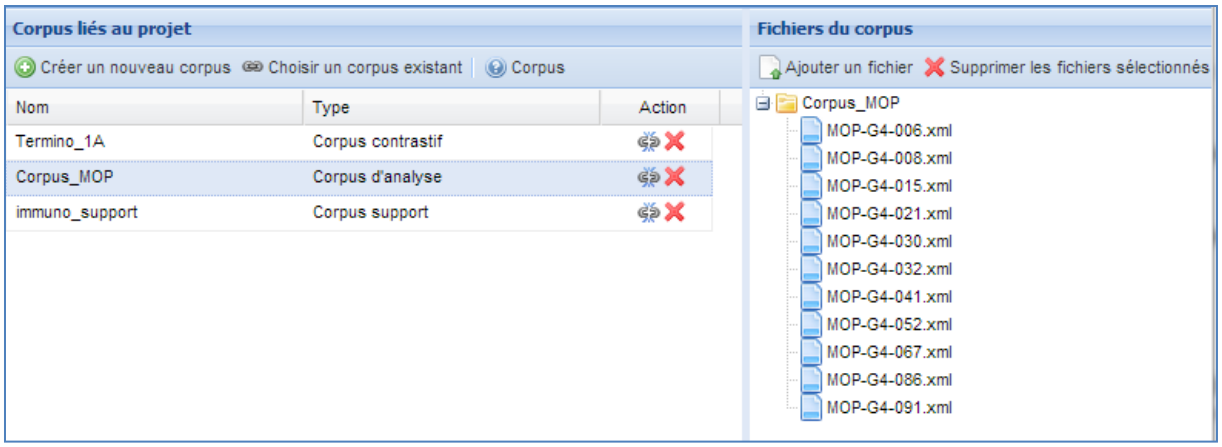


Figure 4-9. Onglet Corpus liés au projet

Si vous êtes Administrateur du Groupe de travail ou Responsable du projet, vous pouvez :

Ajouter un nouveau corpus au projet *Actions utilisateurs*

> cliquer sur *Créer un nouveau corpus* > remplir les champs demandés > cliquer sur *Ajouter ce corpus* ;

➔ le nom du corpus apparait dans la liste;

> sélectionner le nom du corpus dans la liste ;

> dans la fenêtre *Fichiers du corpus*, cliquer sur *Ajouter un fichier* > sélectionner le fichier à télécharger > cliquer sur *Charger ce fichier* ;

> répéter cette opération autant de fois que nécessaire.

Lier un corpus existant au projet *Actions utilisateurs*

> dans la fenêtre *Corpus liés au projet*, cliquer sur *Choisir un corpus existant* > choisir le corpus dans la liste déroulante > choisir le *Type de corpus* dans la liste déroulante > cliquer sur *Lier ce corpus* ;

➔ le corpus apparait dans la liste.

Une analyse ne peut contenir qu'un seul corpus de chaque type. Seul le Corpus d'Analyse est obligatoire.

Retirer un corpus du projet *Actions utilisateurs*

> cliquer sur l'icône à côté du corpus à retirer > confirmer le retrait du corpus.

Supprimer un corpus de la Station

> cliquer sur l'icône à côté du corpus à supprimer > confirmer la suppression du corpus.

Seuls les corpus propriétaires peuvent être supprimés de la Station.

Supprimer un fichier du corpus *Actions utilisateurs*

> dans la fenêtre *Fichiers du corpus*, sélectionner les fichiers à supprimer dans la liste (maintenir la touche CTRL enfoncée durant la sélection de plusieurs fichiers) > cliquer sur *Supprimer les fichiers sélectionnés* > confirmer la suppression.

4.6 Lancer l'analyse (Onglet *Analyse*)

L'onglet *Analyse* est divisé en 3 fenêtres : *Corpus*, *Etat* et *Edition* (cf. Figure 4-10).

Figure 4-10. Onglet *Analyse*

La fenêtre *Corpus* affiche le(s) corpus attaché(s) au Projet, ainsi que l'état de leur analyse. L'analyse jugée intègre par le système est signalée par l'icône , tandis que l'analyse erronée est signalée par l'icône .

Une analyse erronée résulte le plus souvent des erreurs de conversion des fichiers (cf. Préparation du corpus).

Dans le cas d'une analyse erronée d'un ou de plusieurs corpus, il est nécessaire de :

- réviser manuellement le(s) corpus en question (cf. Nettoyage manuel des fichiers);
- re-intégrer le(s) corpus révisé(s) à la Station (cf. Définir et intégrer les corpus);
- relancer l'analyse (cf. Lancer l'analyse).

La fenêtre *Etat* permet de vérifier l'intégrité de l'analyse, une fois celle-ci effectuée. Les résultats du test d'intégrité sont affichés dans la fenêtre *Corpus* (cf. paragraphe précédent).

La fenêtre *Etat* permet aussi de vérifier la disponibilité des ressources externes, qui sont consultées en temps réel, par les Web services. Si les ressources externes sont disponibles, l'icône apparait à côté du nom de chaque ressource. Dans le cas contraire, l'indisponibilité de la ressource est signalée par l'icône .

La fenêtre *Edition* permet de lancer ou d'effacer l'analyse du projet en cours, ainsi que d'accéder à l'espace de travail du projet en cours.

Si vous êtes Administrateur du Groupe de travail ou Responsable du projet, vous pouvez :

Vérifier la disponibilité des ressources externes *Actions utilisateurs*

> dans la fenêtre *Etat*, cliquer sur *Vérifier la disponibilité des ressources externes*.

Vérifier l'intégrité de l'analyse

> dans la fenêtre *Etat*, cliquer sur *Intégrité de l'analyse*.

Effacer l'analyse

> dans la fenêtre *Edition*, cliquer sur *Effacer l'analyse*.

Lancer l'analyse

> dans la fenêtre *Edition*, cliquer sur *Lancer l'analyse*.

Avertissement

Une analyse peut durer très longtemps (plusieurs heures) en fonction de la taille du corpus.

Ouvrir le projet

> dans la fenêtre *Edition*, cliquer sur *Ouvrir le projet*.

Pour accéder à l'espace de travail d'un projet sélectionné, il faut que le projet ait déjà été analysé auparavant.

4.7 Accéder à l'espace de travail d'un projet en cours (Onglet Analyse)

Accéder à l'espace de travail d'un projet en cours *Actions utilisateurs*

> sélectionner le projet dans la fenêtre *Liste des projets* ;
> dans l'onglet *Analyse*, fenêtre *Edition*, cliquer sur *Ouvrir le projet*.

Pour accéder à l'espace de travail d'un projet sélectionné, il faut que le projet ait déjà été analysé auparavant.

5 Analyse automatique

L'analyse automatique permet d'annoter linguistiquement le Corpus d'Analyse par incorporation des résultats des outils et ressources sélectionnés par l'utilisateur. Elle a pour objectif d'extraire de ce corpus des UL (par multi-extraction), de les décrire (résultat des extracteurs et de l'interrogation des ressources définies) et de les pondérer (résultat de l'algorithme de pondération de la Station).

Les informations calculées sont affichées soit dans les fiches lexicales des UL/ SL (cf. Visualisation de fiches lexicales d'UL/SL), soit dans les fiches de relation des UL (cf. Fonctionnalité 'Gérer les relations').

Les informations issues de l'analyse automatique sont, pour chaque UL :

- **Forme Canonique (FC)** : correspond à la forme d'UL trouvée en corpus, priorisée dans l'ordre suivant : TermoStat, YaTea, Acabit.

  ***Note** : Normalement, la Forme Canonique devrait être la forme la plus simple d'une UL, utilisée, par exemple, comme entrée dans les dictionnaires. Cependant, aucun extracteur*

ne fournit de Forme Canonique définie de cette façon; d'où le choix d'utiliser comme Forme Canonique la forme trouvée en corpus par les extracteurs.

- **Forme Lemmatisée (FL)** : correspond, la plupart de temps, à la suite de lemmes de chaque élément d'une UL, ex. *membrane cellulaire* pour l'occurrence *membranes cellulaires* trouvée en corpus ; si plusieurs outils différents proposent une Forme Lemmatisée, une seule Forme Lemmatisée est retenue, dans l'ordre préférentiel suivant : TermoStat, YaTeA, Acabit.
- **Statut lexical (SLEX)** : terminologique ou non, selon le seuil du PT paramétré par l'analyste (cf. Paramétrer les pondérations) ;
- **Domaine(s)** : uniquement si le statut est terminologique ; correspond dans ce cas au domaine renseigné par l'analyste dans le descriptif du projet ; ex. immunobiologie ;
- **Usage** : *préconisé* ou *interdit*, selon les spécifications d'une LC, établies par l'analyste ; par défaut, avant l'analyse manuelle, l'usage préconisé est retenu pour chaque UL.
- **Catégorie(s) sémantique(s) (CAT SEM)** : proposée(s) par les ressources externes (cf. Choisir les ressources externes) ;
- **Fréquence (FREQ)** : nombre d'occurrences des formes fléchies de l'UL en corpus, calculée par la Station Sensunique ;
- **Indices de confiance** :
    - Pondérations internes : **PT, PSL, PUL** (cf. Paramétrer les pondérations);
    - Indices des extracteurs externes : indices de confiance fournis par les extracteurs, ex. loglike pour Acabit (cf. Contenu de la fiche lexicale d'une UL);
- **Tête** : régisseur syntaxique d'une UL, ex. *membrane* pour *membrane cellulaire* ;
- **Expansion** : complément/modifieur d'une Tête, ex. *cellulaire* pour *membrane cellulaire* ;

    ***Note** : Le découpage en Tête et Expansion est fait par les extracteurs ; par conséquent, plusieurs paires de Tête et Expansion peuvent être proposées suite à l'analyse automatique.*

- **Catégorie morphosyntaxique fonctionnelle (CAT FONCT)** : en général, catégorie de la Tête d'une UL, ex. NOM pour *membrane cellulaire ;*
- **Matrice morphosyntaxique (MATRICE MS)** : suite des catégories morphosyntaxiques de chaque élément de l'UL., ex. Nom Adj pour *membrane cellulaire* ;
- **Formes fléchies (FF)** : formes fléchies d'une UL trouvées en corpus, assorties des traits morphosyntaxiques (si possible) et de la fréquence ;
- **Variantes** : provenant soit du corpus analysé, soit des ressources externes, ex. *membrane plasmique* pour *membrane cellulaire* (cf. Fonctionnalité 'Gérer les relations') ;
- **UL dérivées** : UL dont un des composants appartient à la même famille dérivationnelle, ex. *membrane cellulaire* et *marquage de cellule* (cf. Fonctionnalité 'Gérer les relations') ;
- **UL homonymes** : UL homographes d'une autre catégorie morphosyntaxique que l'UL analysée, ex. *contrôle, Nom* et *contrôle, Ver* (cf. Fonctionnalité 'Gérer les relations') ;
- **Collocations (UL liées)** (cf. Fonctionnalité 'Gérer les relations') ;
    - UL incluses : une UL incluse est une UL dont l'intégralité se retrouve dans l'UL analysée ; par exemple, pour l'UL *anticorps monoclonal de souris*, les UL incluses sont : *anticorps monoclonal, anticorps* ;
    - UL composées : une UL composée est une ULC contenant plus que l'intégralité de

l'UL analysée ; par exemple pour l'UL *anticorps monoclonal*, les UL composées sont *anticorps monoclonal conjugué, anticorps monoclonal de souris, anticorps monoclonal HLA-B27* [5];

- UL associées : une UL associée est une UL non incluse et non composée contenant un même lemme que l'UL analysée ; exemple : pour l'UL *anticorps monoclonaux*, UL associée est *solution d'anticorps* ;

- **Sources** :
  - Outil(s) ayant proposé une ULC (exemple : Termostat, Acabit) ;
  - Ressource(s) externe(s) l'attestant (exemple : TermSciences) ;
- **Définition(s)** : les définitions sont extraites des ressources externes consultées (IATE, TermSciences).

Partant du principe que chaque proposition faite lors d'une analyse automatique peut être modifiée, tous les résultats de l'analyse automatique (excepté les indices de confiance calculés par les extracteurs et les sources) sont éditables dans l'Interface de travail de la Station Sensunique.

6 Interface de travail

Les résultats de l'analyse automatique sont affichés dans l'Interface de travail de la Station Sensunique (cf. Figure 6-1). L'Interface de travail rassemble des fonctionnalités facilitant la seconde phase du processus d'acquisition des ressources, à savoir l'analyse manuelle approfondie. Elle consiste en un premier filtrage des UL par un analyste et en l'établissement du consensus final avec les experts métier. Le parti pris fondamental de la Station est que l'analyste peut effectuer tout changement nécessaire concernant l'ensemble des résultats proposés par l'analyse automatique. L'interface de travail lui sert à visualiser, à approfondir et à élargir (si besoin) les résultats afin de les valider pour construire la ressource finale.

L'interface de travail de la Station Sensunique est divisée en 4 espaces (cf. Figure 6-1) :

Espace 1 : Visualisation de la liste d'UL/SL

- Permet de visualiser les résultats d'analyse automatique sous la forme d'une liste d'UL/ SL avec leurs champs associés (cf. Onglet Liste d'UL et Onglet Liste de SL) ;
- Permet de gérer les relations entre les différentes UL (cf. Fonctionnalité Gérer les relations) ;
- Permet de visualiser les résultats de la recherche en corpus (cf. Onglet Résultats de la recherche) ;
- Permet de visualiser les résultats de l'analyse (cf. Onglet Statistiques de l'analyse).

Espace 2 : Filtres sur la liste d'UL/SL

- Permet de filtrer les résultats sur la liste d'UL/ SL.

Espace 3 : Visualisation d'UL en contexte

- Permet de visualiser les UL sélectionnées en corpus;
- Permet de visualiser les résultats de la recherche en corpus.

Espace 4 : Visualisation de fiches lexicales d'UL/SL

- Permet de visualiser et modifier les fiches d'UL/ SL.

[5] Les UL incluses et composées fonctionnent de manière symétrique : si l'ULC1 est ULC incluse d'une ULC2, alors l'ULC2 sera ULC composée de l'ULC1.

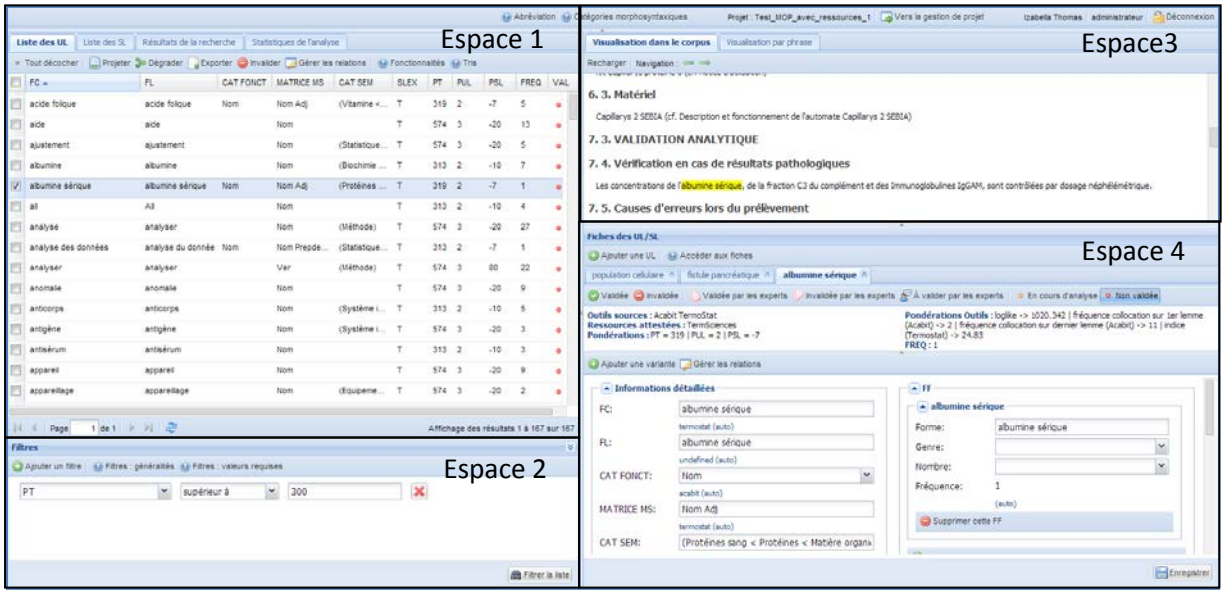


Figure 6-1. Interface de travail

Un cinquième espace *Recherche/filtres* (cf. Concordancier évolué) peut être ouvert dans l'interface de travail. Il permet d'effectuer des recherches sur le Corpus d'Analyse à l'aide d'un concordancier évolué (cf. Figure 6-2).

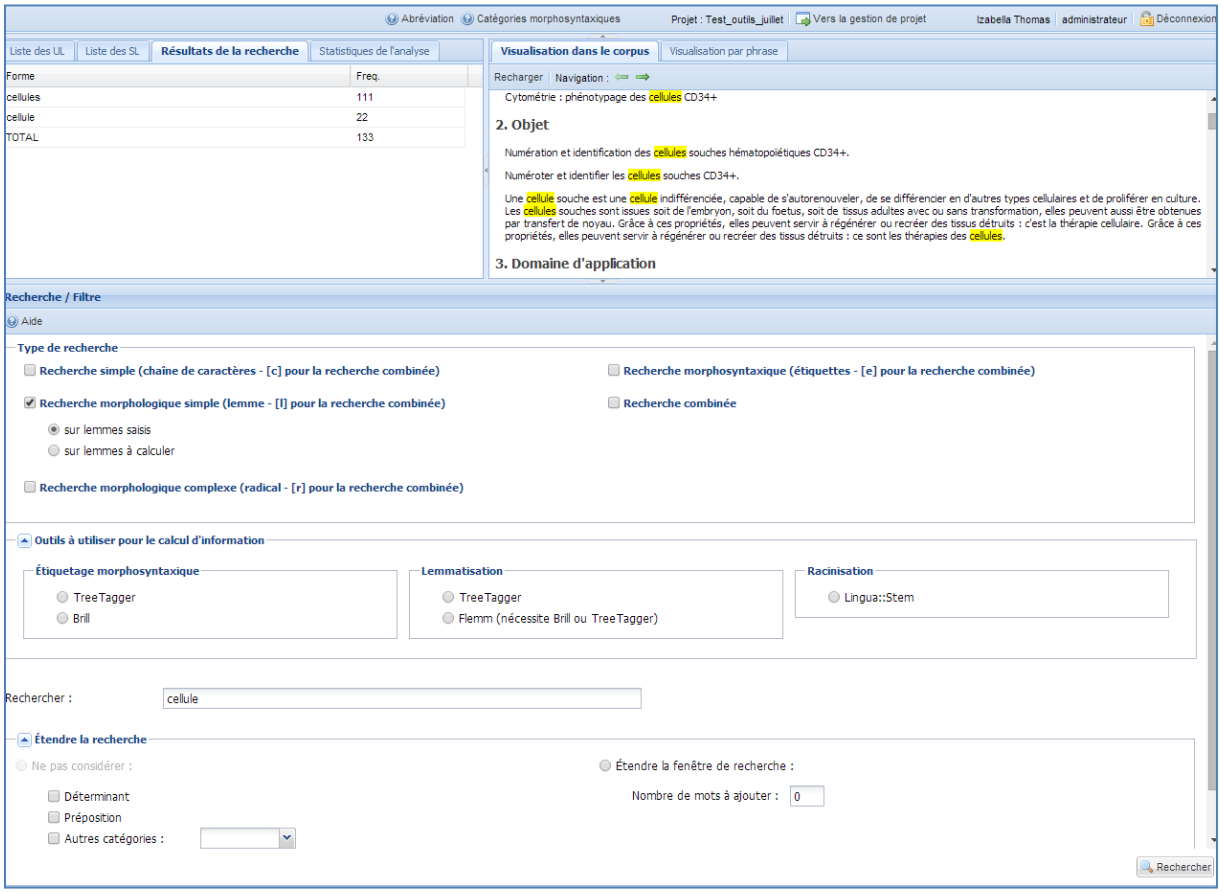


Figure 6-2. Concordancier évolué

L'espace de travail est modifiable. Chaque fenêtre peut être redimensionnée, affichée ou cachée.

Redimensionner une fenêtre *Actions utilisateurs*

> positionner le pointeur de la souris sur la bordure de la fenêtre à redimensionner > faire glisser la bordure dans la direction souhaitée en maintenant enfoncé le bouton gauche de la souris.

Afficher ou cacher une fenêtre

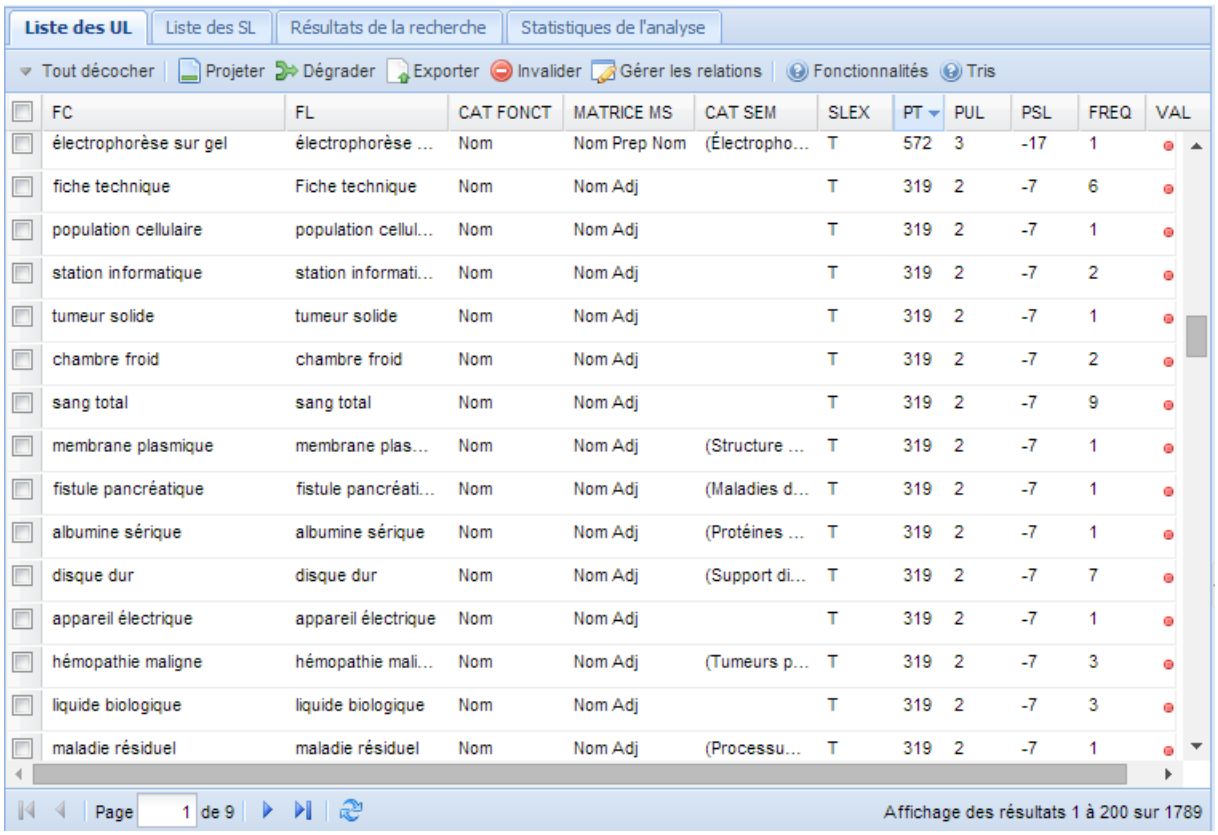
> positionner le pointeur de la souris sur la flèche sur la bordure de la fenêtre à afficher / cacher (cf. Figure 6-3) > effectuer un clic gauche pour afficher/ cacher la fenêtre.

Figure 6-3. Flèche sur la bordure de la fenêtre à afficher/cacher

6.1 Espace 1 : Visualisation de la liste d'UL/ SL

6.1.1 Onglet Liste des UL

L'onglet *Liste des UL* permet tout d'abord d'afficher les UL proposées par les différents outils, sous format d'une liste, avec des champs contenant les informations associées à chaque UL (cf. Figure 6-4), à savoir Forme Canonique (FC), Forme Lemmatisée (FL), Catégorie Fonctionnelle (CAT FONC), Matrice Morphosyntaxique (MATRICE MS), Catégorie Sémantique (CAT SEM), Statut Lexical (SLEX), Poids Terminologique (PT), Poids d'Unité Lexicale (PUL), Poids de Structure Lexicale(PSL), Fréquence (FREQ), Statut de Validation (VAL) (cf. Analyse Automatique).



FC	FL	CAT FONCT	MATRICE MS	CAT SEM	SLEX	PT	PUL	PSL	FREQ	VAL
électrophorèse sur gel	électrophorèse ...	Nom	Nom Prep Nom	(Électropho...	T	572	3	-17	1	
fiche technique	Fiche technique	Nom	Nom Adj		T	319	2	-7	6	
population cellulaire	population cellul...	Nom	Nom Adj		T	319	2	-7	1	
station informatique	station informati...	Nom	Nom Adj		T	319	2	-7	2	
tumeur solide	tumeur solide	Nom	Nom Adj		T	319	2	-7	1	
chambre froid	chambre froid	Nom	Nom Adj		T	319	2	-7	2	
sang total	sang total	Nom	Nom Adj		T	319	2	-7	9	
membrane plasmique	membrane plas...	Nom	Nom Adj	(Structure ...	T	319	2	-7	1	
fistule pancréatique	fistule pancréati...	Nom	Nom Adj	(Maladies d...	T	319	2	-7	1	
albumine sérique	albumine sérique	Nom	Nom Adj	(Protéines ...	T	319	2	-7	1	
disque dur	disque dur	Nom	Nom Adj	(Support di...	T	319	2	-7	7	
appareil électrique	appareil électrique	Nom	Nom Adj		T	319	2	-7	1	
hémopathie maligne	hémopathie mali...	Nom	Nom Adj	(Tumeurs p...	T	319	2	-7	3	
liquide biologique	liquide biologique	Nom	Nom Adj		T	319	2	-7	3	
maladie résiduel	maladie résiduel	Nom	Nom Adj	(Processu...	T	319	2	-7	1	

Figure 6-4. Onglet Liste des UL

Différentes actions peuvent être effectuées sur la liste d'UL.

Trier la liste des UL par ordre croissant / décroissant *Actions utilisateurs*

> cliquer sur la flèche qui apparait dans le titre de la colonne correspondant à la propriété souhaitée pour le tri.

Sélectionner / désélectionner une ou plusieurs UL

> cocher/ décocher précédant l'UL ;

➔ l'UL est sélectionnée .

➔ l'UL est désélectionnée .

Sélectionner/ désélectionner toutes les UL

> Cocher/ décocher précédant la FC ;

➔ FC : Toutes les UL de la liste sont désélectionnées.

➔ FC : Toutes les UL de la liste sont sélectionnées.

Afficher la fiche lexicale d'une UL

> dans l'espace *Liste des UL* double-cliquer sur l'UL à afficher ;

➔ la fiche lexicale de l'UL sélectionnée s'affiche dans l'espace *Fiches des UL/ SL*.

Par ailleurs, l'utilisateur peut effectuer plusieurs actions sur la Liste des UL à partir des fonctionnalités affichées en haut de l'onglet (cf. Figure 6-5).

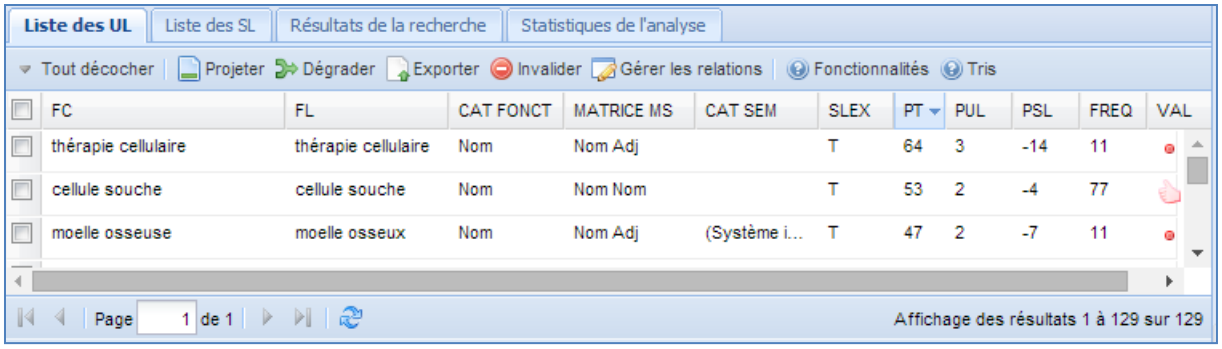

Liste des UL
Liste des SL
Résultats de la recherche
Statistiques de l'analyse
Tout décocher
Projeter
Dégrader
Exporter
Invalider
Gérer les relations
Fonctionnalités
Tris
FC FL CAT FONCT MATRICE MS CAT SEM SLEX PT PUL PSL FREQ VAL
thérapie cellulaire thérapie cellulaire Nom Nom Adj T 64 3 -14 11
cellule souche cellule souche Nom Nom Nom T 53 2 -4 77
moelle osseuse moelle osseux Nom Nom Adj (Système i... T 47 2 -7 11
Page 1 de 1
Affichage des résultats 1 à 129 sur 129

Figure 6-5. Fonctionnalités dans l'Onglet Liste des UL

6.1.1.1 Fonctionnalité Tout décocher Tout décocher

Cette fonctionnalité permet de décocher toutes les UL sélectionnées au préalable.

Décocher toutes les UL sélectionnées *Actions utilisateurs*

> cliquer sur le bouton Tout décocher ;

➔ toutes les UL sont désélectionnées.

6.1.1.2 Fonctionnalité Projeter

Cette fonctionnalité sert à visualiser en corpus les UL sélectionnées dans la liste des UL (cf. Figure 6-6).

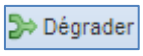
Visualiser une ou plusieurs UL en corpus ou par phrase *Actions utilisateurs*

> sélectionner une ou plusieurs UL dans la liste des UL > cliquer sur le bouton

> aller dans l'espace *Visualisation dans le corpus / Visualisation par phrase* ;

➔ l'UL ou les UL sélectionnée(s) apparai(ssen)t surlignée(s) en jaune.

Figure 6-6. Visualisation des UL

6.1.1.3 Fonctionnalité Dégrader 

Cette fonctionnalité permet de transformer une UL ou un groupe d'UL en Structure Lexicale (SL). Nous introduisons la notion de Structure Lexicale pour répondre au critère de non-ambigüité d'une LC tout en conservant le caractère exhaustif du lexique et la nécessité de restriction d'emploi selon le contexte. La notion de Structure Lexicale dépasse la définition d'unité lexicale à strictement parler puisqu'elle s'appuie sur la combinatoire lexico-syntaxique entre plusieurs unités lexicales, se situant ainsi à la frontière du lexique et de la syntaxe. Cette notion est à rapprocher de celles de classes de sélection distributionnelles, classes d'objets, fonctions lexicales, cadres prédicatifs, pour ne citer que quelques-unes des dénominations décrivant ces types de construction dans différentes théories linguistiques.

On définit une Structure Lexicale (SL) comme un patron morphosyntaxique imposé et contrôlé par un lexème, souvent prédicatif, composée d'une partie figée (lexicalisée, variable uniquement en flexion) et d'une partie variable (mais contrainte par des traits morphosyntaxiques et sémantiques). Par exemple, *marquage* est le lexème prédicatif dans *marquage des cellules, marquage des cellules leucocytaires, marquage des cellules endothéliales vasculaires animales, marquage des cellules en suspension*. Le besoin de définir des structures lexicales vient, d'une part, de l'impossibilité d'encoder ces constructions dans un dictionnaire de termes (puisque ce ne sont pas des UL) et, d'autre part, de la nécessité de contrôler leur distribution et leur variabilité dans un environnement de rédaction d'une LC. C'est pour ces raisons que nous proposons de les recenser dans un dictionnaire spécifique, sous un format décrivant leurs principales caractéristiques :

Exemple
marquage de < NOM : CELLULE>

La partie variable, introduite par les chevrons (<>), est généralement définie par sa catégorie fonctionnelle (ici : NOM), qui peut être en plus caractérisée par son appartenance à une classe sémantique (ici : CELLULE).

La notion de Structure Lexicale est primordiale lorsque, nous éloignant de la théorie terminologique classique, nous considérons comme termes des syntagmes autres que les syntagmes nominaux. En effet, des verbes ou des adjectifs peuvent renvoyer à des concepts bien spécifiques dans des domaines précis. Certains dictionnaires terminologiques recensent d'ores et déjà des termes de nature verbale. Par exemple, on trouve aussi bien le nom 'centrifugation' que le verbe 'centrifuger' dans Le Grand Dictionnaire Terminologique. Simplement, la description de ce verbe, en s'arrêtant à l'identification de sa catégorie verbale, ne nous renseigne ni sur la présence ni sur la nature de ses compléments : pourtant, on *centrifuge* toujours quelque chose, *du sang total, du plasma sanguin* etc. Nous proposons donc de recenser ce verbe dans un dictionnaire de structures, en indiquant clairement qu'il doit être accompagné de compléments d'une certaine classe fonctionnelle et sémantique : centrifuger <NOM : SANG>.

Un autre avantage concernant l'identification des structures lexicales est l'établissement des relations entre des UL dérivées et la vérification de la cohérence du recensement du vocabulaire. En théorie, les UL prédicatives en relation de dérivation ne peuvent introduire dans leurs structures que des compléments appartenant à des classes sémantiques identiques :

Exemple
numéroter < NOM : CELLULE> ; <NOM : CELLULE> numéroté(es) ; numération de <NOM : CELLULE>
Pour rédiger *: numération des populations leucocytaires*, *numéroter les lymphocytes T, B et NK*

L'avantage du recensement de ces structures est double : d'une part, cela permet de contrôler que *populations leucocytaires* et *lymphocytes T, B et NK* portent bien la contrainte sémantique *CELLULE* et que *numéroter, numération* (voir le participe passé adjectival *numéroté*) renvoient toujours à la même classe sémantique.

La dégradation d'une UL ou d'un groupe d'UL en SL se fait manuellement, aucune proposition automatique n'est fournie par cette fonctionnalité. Par ailleurs, la formalisation des SL que nous proposons n'est qu'indicative et peut être redéfinie en fonction du projet et des besoins spécifiques.

Transformer une UL ou un groupe d'UL en une SL *Actions utilisateurs*

> sélectionner les UL appropriées dans la liste des UL > cliquer sur le bouton *Dégrader* ;

➔ la fenêtre *Création d'une nouvelle structure lexicale* apparait (cf. Figure 6-7) ;

> remplir les informations demandées > cocher la/les UL à conserver dans la liste des UL > valider ;

➔ la nouvelle SL apparait dans la liste des SL.

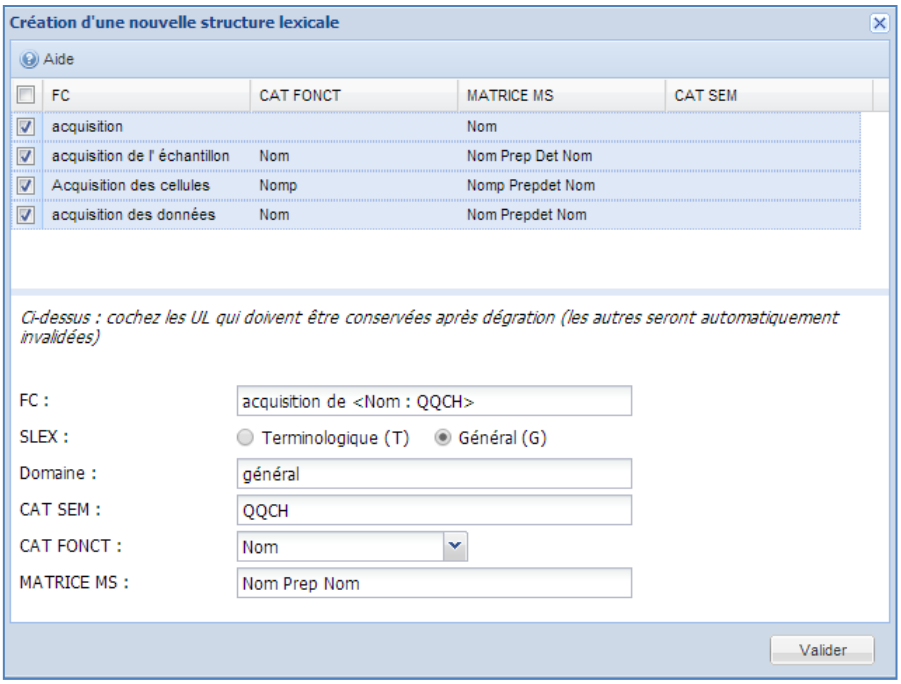
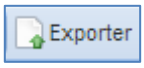


Figure 6-7. Création d'une SL

Une UL ayant servi à la création d'une SL, peut garder son statut de l'UL et, par conséquent, elle peut être conservée dans la liste des UL. Dans ce cas la(les) SL issue(s) de l'UL conservée est/sont indiquée(s) dans sa fiche lexicale (cf. Contenu de la fiche lexicale d'une UL).

6.1.1.4 *Fonctionnalité Exporter (UL)* 

Cette fonctionnalité permet d'exporter en dictionnaires les données recensées dans la Station au format XML afin de :

- créer des ressources terminologiques diverses ;
- exploiter les données dans d'autres applications ;
- valider, durant l'analyse, les données nécessitant des compétences spécifiques par des experts métiers.

En fonction de son objectif, l'utilisateur peut paramétrer les dictionnaires de sortie, en choisissant le(s) type(s) d'informations qu'il souhaite exporter (cf. Figure 6-8). Toute la finesse de description d'une ressource produite dans la Station n'est pas forcément utile à l'application qui va exploiter cette ressource. De même, on peut n'être intéressé que par un périmètre restreint des UL recensées.

La sélection s'effectue à l'aide des filtres cumulatifs servant à restreindre le périmètre des données exportées selon deux axes :

- sélection des propriétés des UL (parmi les 17 propriétés proposées, telles que définition, synonymes, matrice morphosyntaxique, catégorie sémantique, colocations, statut de validation, etc.) :

  Exemple : dictionnaire d'UL contenant seulement : Forme canonique, Définition et Variantes
  Exemple : dictionnaire d'UL contenant seulement : Forme canonique, Matrice morphosyntaxique et Fréquence

- sélection des propriétés des UL et des valeurs de propriétés :

  Exemple : dictionnaire d'UL contenant seulement : Forme canonique, Classe Sémantique, Définition, Statut de Validation ; ET le Statut de Validation est *Validée*

Le même projet permet de créer plusieurs ressources en fonction d'une application visée. Le principe est le même pour les dictionnaires de SL (cf. Fonctionnalité Exporter (SL)).

Actions utilisateurs

Créer et exporter des dictionnaires

> cliquer sur le bouton *Exporter* ;

➔ la fenêtre *Export du dictionnaire d'UL* apparait ;

> cocher les champs à exporter > appliquer des filtres si vous le souhaitez > cliquer sur le bouton *Exporter* ;

➔ les fichiers sont exportés au format XML.

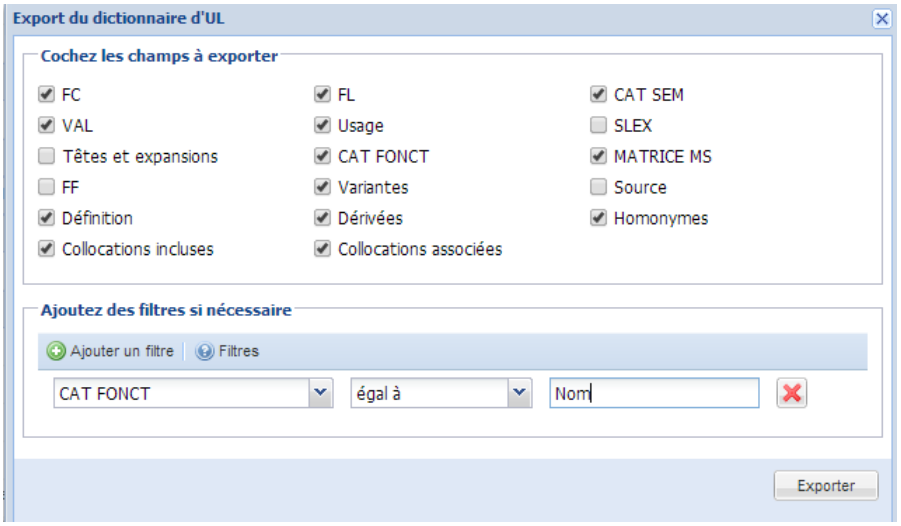


Figure 6-8. Export du dictionnaire d'UL

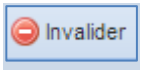
Les filtres sur l'export fonctionnent de la même manière que les filtres sur la liste des UL (cf. Filtres sur la liste des UL/ SL).

- **Filtres sur les chaînes de caractères** :
  Champs à filtrer : FC, FF, FL, CAT FONCT, MATRICE MS, CAT SEM, Variante (FC et FF), Tête, Expansion, Définition, SLEX, Domaine, Type d'usage, Communauté d'usage, Outil, Ressource
  Opérateurs *: égal à, différent de*
  Valeurs : expressions régulières de type PCRE entourées de '/'
  Exemples : /cellule souche/, /Nom Nom/, /^cellule/, /^cellule$/ etc.
  Pour en savoir plus sur PCRE : http://fr2.php.net/manual/fr/book.pcre.php

- **Filtres sur les entiers** :
  Champs à filtrer : PT, PUL, PSL, FREQ, VAL
  Opérateurs *: égal à, différent de, supérieur à, supérieur ou égal à, inferieur à, inferieur ou égal à*
  Valeurs : nombres entiers
  Exemples : 3, 22, 145, etc.

Certains filtres demandent des **valeurs spécifiques**, en fonction de la configuration de la Station :

- **SLEX** : /G/ ('Général') ou /T/ ('Terminologique')
- **Type d'usage** : /préconisé/, /interdit/
- **Communauté d'usage** : /professionnel/, /public/ (pour 'Grand public')
- **VAL** : 0 ('Non validée')
  1 ('En cours d'analyse')
  2 ('A valider par les experts')
  3 ('Invalidée par les experts')
  4 ('Validée par les experts')
  5 ('Invalidée')
  6 ('Validée')
- **Outils** : /acabit/, /YaTea/, /termostat/
- **Ressources** : /iate/, /termsciences/

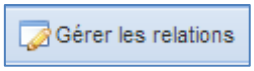
6.1.1.5 *Fonctionnalité Invalider* 

Invalider une ou plusieurs UL de la liste des UL *Actions utilisateurs*

> sélectionner une ou plusieurs UL dans la liste des UL > cliquer sur le bouton *Invalider* ;

➔ un message d'avertissement apparait pour vous prévenir que cette action entraîne la suppression totale des UL sélectionnées et de leurs associations avec les autres UL ;

> valider/invalider.

6.1.1.6 *Fonctionnalité Gérer les relations* 

Cette fonctionnalité permet de visualiser et d'établir de nouvelles relations entre les UL, telles que :

- relations morphologiques :
  - **FF (Formes Fléchies)** : regroupe les formes fléchies de l'UL analysée ;

  Exemple : pour l'UL *cellule*, les FF sont *cellule*, *cellules* ;

  - **UL dérivées** : regroupe les UL dont un des composants appartient à la même famille dérivationnelle ;

  Exemple : pour l'UL membrane **cellulaire**, l'UL dérivée est marquage de **cellules** ;

- relations lexico-syntaxiques :
  - **UL incluses** : une UL incluse est une UL dont l'intégralité se retrouve dans l'UL analysée ;

  Exemple : pour l'UL anticorps monoclonal de souris, les UL incluses sont : anticorps monoclonal, anticorps, monoclonal etc. ;

  - **UL composées** : une UL composée est une UL contenant plus que l'intégralité de la ULC analysée ;

  Exemple : pour l'UL anticorps monoclonal, les UL composées sont anticorps monoclonal conjugué, anticorps monoclonal de souris, anticorps monoclonal HLA-B27 etc.6;

  - **UL associées** : une UL associée est une UL non incluse et non composée contenant un même lemme que l'UL analysée ;

[6] Les UL incluses et composées fonctionnent de manière symétrique : si une ULC1 est ULC incluse d'une ULC2, alors l'ULC2 sera ULC composée de l'ULC1.

Exemple : pour l'UL anticorps monoclonaux, UL associée est solution d'anticorps, anticorps primaires, etc. ;

- relations lexico-sémantiques :
  - **UL variantes** : regroupe les synonymes (abréviation, acronyme, variante morphosyntaxique ou orthographique) de l'UL analysée, provenant soit du corpus analysé, soit des ressources externes ;

Exemple : pour l'UL anticorps monoclonal, les variantes sont Ac Mo, AcM, ACMC ;
Exemple : pour l'UL marquage cellulaire, la variante est marquage de cellules.

  - **UL homonymes :** une UL homonyme est une UL formellement identique mais de catégorie morphosyntaxique (CAT MS) différente que l'UL analysée.

Exemple : pour l'UL *étude (Nom)*, l'UL homonyme est *étude (Verbe)*.

Actions utilisateurs

Ajouter/supprimer les relations entre les UL (cf. Figure 7-9)

> sélectionner une ou plusieurs UL dans la liste des UL > cliquer sur le bouton *Gérer les relations ;*

➔ La fenêtre *Réseau de relations de l'UL : nom de l'UL* apparait ;

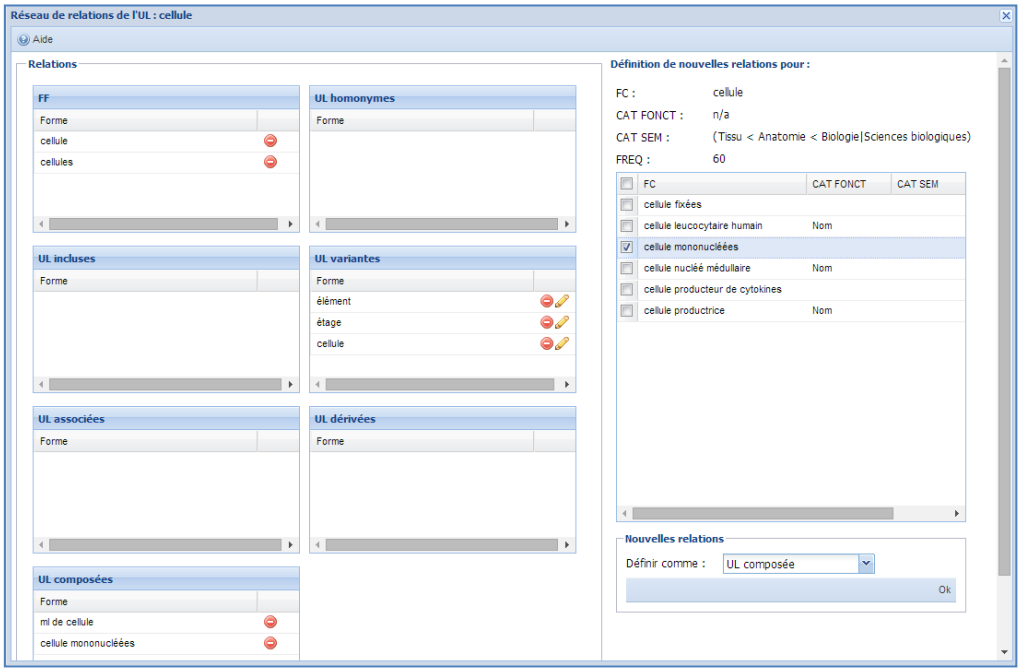


Figure 6-9. Fiche de relation d'une UL

La fenêtre *Réseau de relations de l'UL : nom de l'UL* est subdivisée en 2 espaces : *Relations* et *Définir de nouvelles relations pour* (une UL).

Dans l'espace *Relations* sont affichées toutes les relations de l'UL analysée avec d'autres UL, calculées automatiquement par la Station ou définies par l'analyste.

Dans l'espace *Définir de nouvelles relations pour* l'analyste peut définir de nouvelles relations entre les UL ou supprimer les relations existantes.

Actions utilisateurs

Définir une nouvelle relation entre l'UL analysée et le(s) UL présélectionnée(s)

> sélectionner une ou plusieurs UL dans l'espace *Définir de nouvelles relations pour* > choisir le type de relation dans le menu déroulant dans l'encart *Définir comme* en bas à droite de la fenêtre > cliquer sur OK ;

➔ l'UL ou les UL apparaissent dans l'encart correspondant au type d'UL choisi dans l'espace *Relations*.

Si la relation sélectionnée est UL variante, il faut aussi sélectionner le *Type de l'UL variante* entre:

- **Forme Abrégée** : il s'agit des acronymes ou des abréviations d'une UL analysée ;

  Exemple : formes abrégées de l'UL *anticorps primaire* : *Ac Iaire* ;
  Exemple : formes abrégées de l'UL *anticorps monoclonal* : *Ac Mo*, *AcM*, *ACMC* ;

- **Synonyme** : il s'agit d'UL répertoriées comme synonymes selon l'utilisateur ou selon une ressource attestée ;

  Exemple : synonyme de l'UL *anticorps monoclonal* : *monoclonal* ;

- **Variante morphologique dérivationnelle** : une variante impliquant une dérivation entre un élément de 2 UL :

  Exemple : variante morphologique dérivationnelle de l'UL *marquage cellulaire* : *marquage de cellules* (dérivation entre *cellule/cellulaire*) ;

- **Forme à variation syntaxique faible** : il s'agit de formes présentant des petits changements de structure, tels que l'insertion ou la variation (au sens changement) d'un mot grammatical ;

  Exemple : forme à variation syntaxique faible de l'UL *cytomètre de flux* : *cytomètre en flux*.

- **Autre Variante** : il s'agit des autres types de variantes, non-recensés dans les types précédents, par exemple des variantes (correctes) orthographiques comme dans *anévrisme / anévrysme*.

Actions utilisateurs

Supprimer une relation existante

> dans l'espace *Relations*, cliquer sur l'icône à côté d'une UL à supprimer ;

➔ l'UL disparait de l'encart.

Editer la fiche d'une UL variante

> dans l'espace *Relations*, cliquer sur l'icône dans l'encart *UL variantes* ;

➔ la fiche lexicale de l'UL variante s'affiche ;

> modifier la fiche si nécessaire > cliquer sur *Modifier*.

Définir les relations pour une UL faisant partie de la sélection initiale (UL se trouvant dans l'encart à droite) :

> dans l'espace *Définir des nouvelles relations pour*, cliquer sur l'icône à côté de l'UL dont on veut définir les relations ;

➔ la nouvelle fenêtre *Réseaux des relations de l'UL* s'affiche.

6.1.2 Onglet Liste des SL

L'onglet *Liste des SL* permet d'afficher les SL issues du processus de dégradation (cf. Fonctionnalité Dégrader), sous format d'une liste, avec des champs contenant les informations associées à chaque SL (cf. Figure 6-10), à savoir Forme Canonique (FC), Catégorie Fonctionnelle (CAT FONC), Matrice Morphosyntaxique (MATRICE MS), Catégorie Sémantique (CAT SEM), Statut Lexical (SLEX), Fréquence (FREQ).

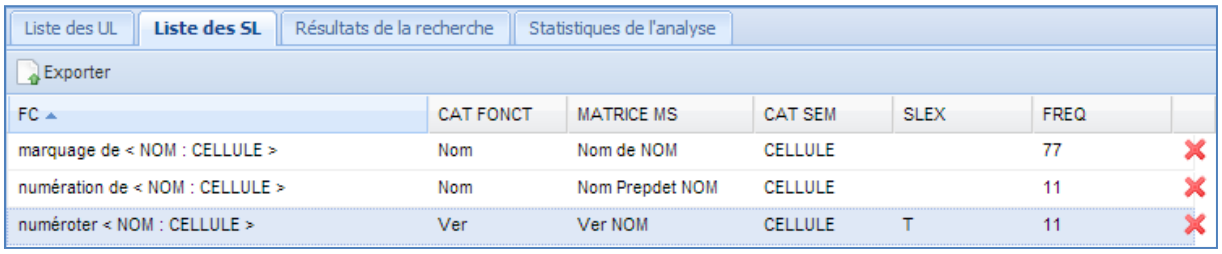


Figure 6-10. Liste des SL

Différentes actions peuvent être effectuées sur la liste de SL.

Actions utilisateurs

Trier la liste des SL par ordre croissant / décroissant

> cliquer sur la flèche qui apparait dans le titre de la colonne correspondant à la propriété souhaitée pour le tri.

Actions utilisateurs

Afficher la fiche lexicale d'une SL

> dans l'espace *Liste des SL* double-cliquer sur la SL à afficher ;

➔ la fiche lexicale de l'UL sélectionnée s'affiche dans l'espace *Fiches des UL/ SL*.

Supprimer une SL da la liste des SL

> cliquer sur l'icône à côté de la SL à supprimer ;

➔ un message demandant une confirmation de la suppression s'affiche ;

> confirmer /infirmer ;

➔ si confirmé, la SL disparait de la Liste des SL.

Par ailleurs, l'utilisateur peut exporter la Liste des SL à partir de la fonctionnalité

affichée en haut de l'onglet.

6.1.2.1 Fonctionnalité Exporter (SL)

Cette fonctionnalité permet d'exporter en dictionnaires les données recensées dans la Station au format XML afin de :

- créer des ressources terminologiques diverses ;
- exploiter les données dans d'autres applications ;
- valider, durant l'analyse, les données nécessitant des compétences spécifiques par des experts métiers.

En fonction de son objectif, l'utilisateur peut paramétrer les dictionnaires de sortie, en choisissant le(s) type(s) d'informations qu'il souhaite exporter (cf. Figure 6-11).

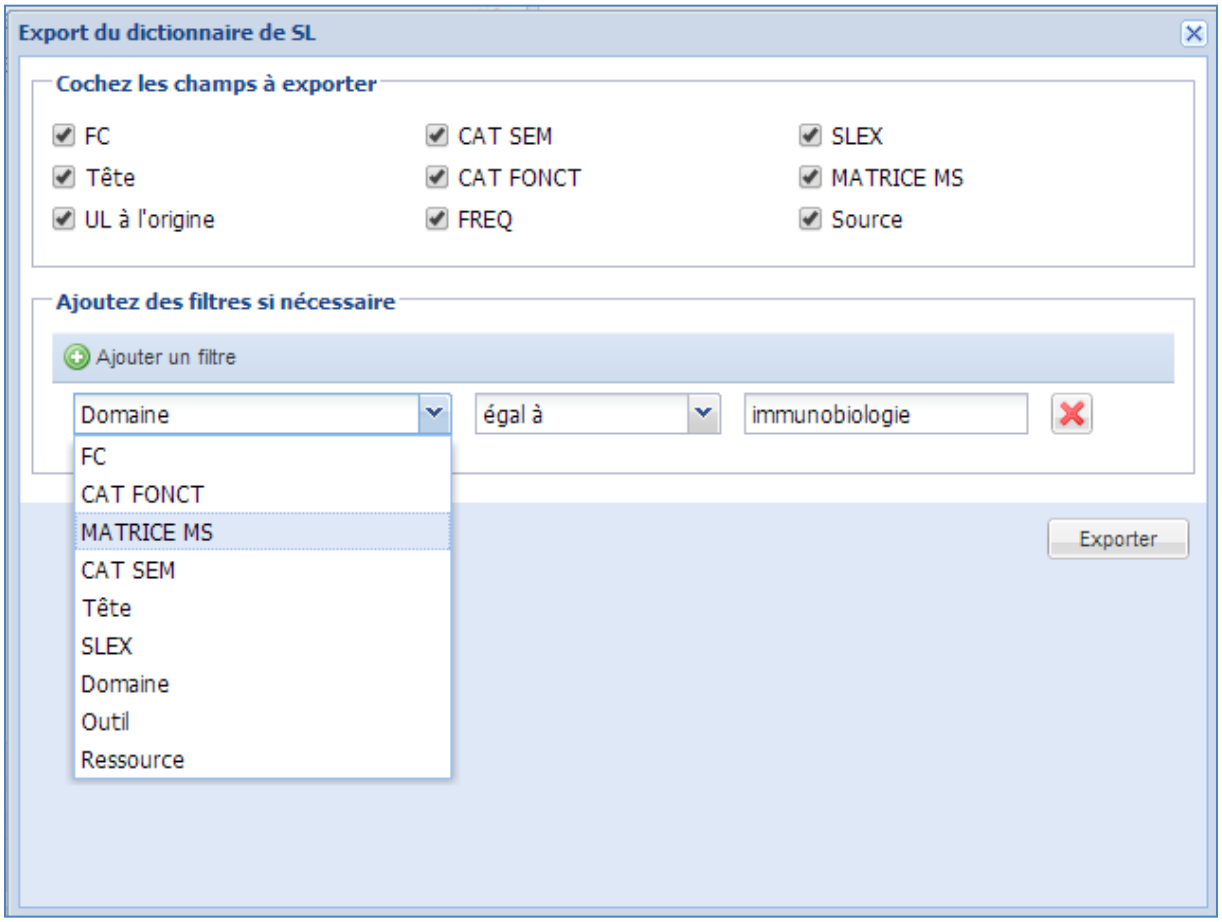


Figure 6-11.Export du dictionnaire d'une SL

Créer et exporter des dictionnaires *Actions utilisateurs*

> cliquer sur le bouton *Exporter* ;

➔ la fenêtre *Export du dictionnaire de SL* apparait ;

> cocher les champs à exporter > appliquer des filtres si vous le souhaitez > cliquer sur le bouton *Exporter* ;

➔ les fichiers sont exportés en format XML.

Les filtres sur l'export fonctionnent de la même manière que les filtres sur la liste des UL (cf. Filtres sur la liste des UL/ SL).

- **Filtres sur les chaînes de caractères** :
  Champs à filtrer : FC, CAT FONCT, MATRICE MS, CAT SEM, Tête, SLEX, Domaine, Outil, Ressource
  Operateurs *: égal à, différent de*
  Valeurs : expressions régulières de type PCRE entourées de '/'
  Exemples : /cellule souche/, /Nom Nom/, /^cellule/, /^cellule$/ etc.
  Pour en savoir plus sur PCRE : http://fr2.php.net/manual/fr/book.pcre.php

Certains filtres demandent des **valeurs spécifiques**, en fonction de la configuration de la station :

- **SLEX** : /G/ ('Général') ou /T/ ('Terminologique')
- **Outils** : /acabit/, /YaTea/, /termostat/
- **Ressources** : /iate/, /termsciences/

6.1.3 Onglet Résultats de la recherche

Cet onglet permet d'afficher les résultats de la recherche sur le corpus, effectuée à l'aide du concordancier (cf. Concordancier évolué). Les résultats sont présentés dans un tableau contenant les formes trouvées et leur fréquence (cf. Figure 6-12).

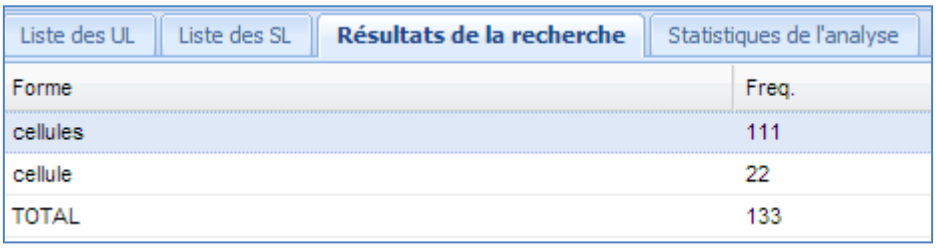


Figure 6-12. Onglet Résultats

Aucune action ne peut être effectuée sur ces résultats (dans cet onglet).

6.1.4 Onglet Statistiques de l'analyse

Cet onglet permet de visualiser les statistiques de l'analyse (cf. Figure 6-13) à chaque moment de son déroulement.

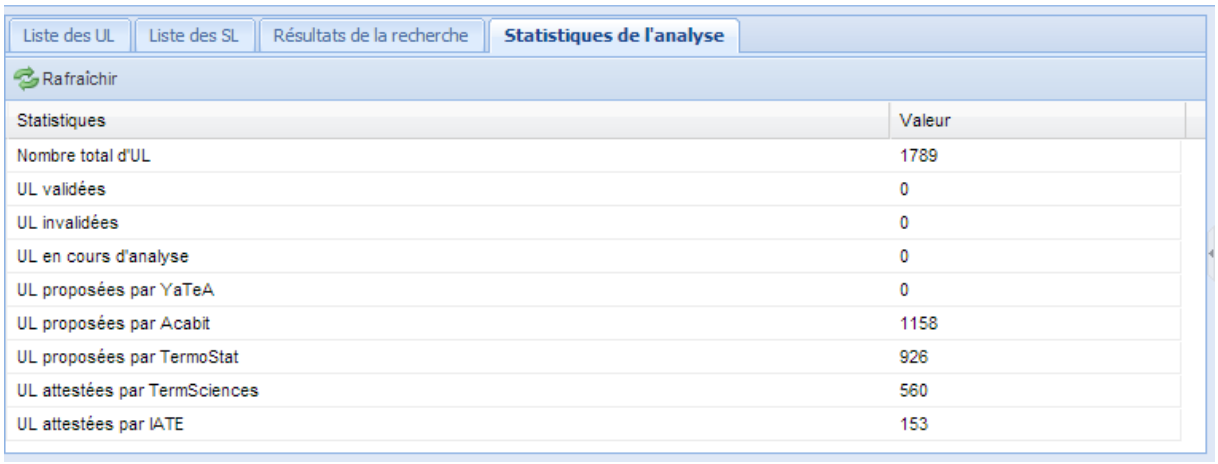


Figure 6-13. Onglet Statistiques de l'analyse

Après chaque modification manuelle dans les résultats de l'analyse (ajout, suppression, validation, invalidation d'une UL), les statistiques peuvent être mises à jour.

Rafraichir les statistiques de l'analyse : *Actions utilisateurs*

> cliquer sur l'icône Rafraîchir ;

➔ les statistiques d'analyse sont mises à jour.

6.2 Espace 2 : Filtres sur la liste des UL/SL

Les filtres permettent d'afficher des UL/ SL respectant les critères choisis par l'utilisateur. Ils sont temporaires et peuvent être modulés à volonté. Les filtres sont cumulés, c'est-à-dire que chaque filtre est calculé indépendamment des autres et que l'intersection des résultats est affichée. On peut ajouter autant de filtres que nécessaire.

Actions utilisateurs

Mettre en place un filtre

> cliquer sur *Ajouter un filtre* > choisir le champ à filtrer dans la première liste déroulante > choisir l'opérateur dans la seconde liste déroulante > entrer la valeur recherchée dans le dernier champ > cliquer sur *Filtrer la liste* ;

➔ les UL filtrées s'affichent dans l'espace *Liste des UL*.

Enlever un filtre

> cliquer sur l'icône à côté du filtre à supprimer > cliquer sur *Filtrer la liste*.

Revenir à la liste initiale

> enlever tous les filtres > cliquer sur *Filtrer la liste*.

6.2.1 Filtres : Valeurs requises

- **Filtres sur les chaînes de caractères** :
  Champs à filtrer : FC, FF, FL, CAT FONCT, MATRICE MS, CAT SEM, Variante (FC et FF), Tête, Expansion, Définition, SLEX, Domaine, Type d'usage, Communauté d'usage, Outil, Ressource
  Opérateurs *: égal à, différent de*
  Valeurs : expressions régulières de type PCRE entourées de '/'
  Exemples : /cellule souche/, /Nom Nom/, /^cellule/, /^cellule$/ etc.
  Pour en savoir plus sur PCRE : http://fr2.php.net/manual/fr/book.pcre.php

- **Filtres sur les entiers** :
  Champs à filtrer : PT, PUL, PSL, FREQ, VAL
  Opérateurs *: égal à, différent de, supérieur à, supérieur ou égal à, inférieur à, inférieur ou égal à*
  Valeurs : nombres entiers
  Exemples : 3, 22, 145, etc.

Certains filtres demandent des **valeurs spécifiques**, en fonction de la configuration de la station :

- **SLEX** : /G/ ('Général') ou /T/ ('Terminologique')
- **Type d'usage** : /préconisé/, /interdit/
- **Communauté d'usage** : /professionnel/, /public/ (pour 'Grand public')
- **VAL** : 0 ('Non validée')
  1 ('En cours d'analyse')
  2 ('A valider par les experts')
  3 ('Invalidée par les experts')
  4 ('Validée par les experts')
  5 ('Invalidée')

6 ('Validée')

- **Outils** : /acabit/, /YaTea/, /termostat/
- **Ressources** : /iate/, /termsciences/

6.3 Espace 3 : Visualisation de fiches lexicales d'UL/SL

Les fiches lexicales des UL/SL contiennent toutes les informations relatives à l'UL/ SL analysée (cf. Figure 6-14). L'analyste peut ajouter, modifier, compléter, valider ou supprimer toute UL/SL ou information la concernant à partir de sa fiche lexicale. Chaque proposition/modification de données est tracée, c'est-à-dire, assortie du nom de son auteur (qu'il soit analyste, outil ou ressource).

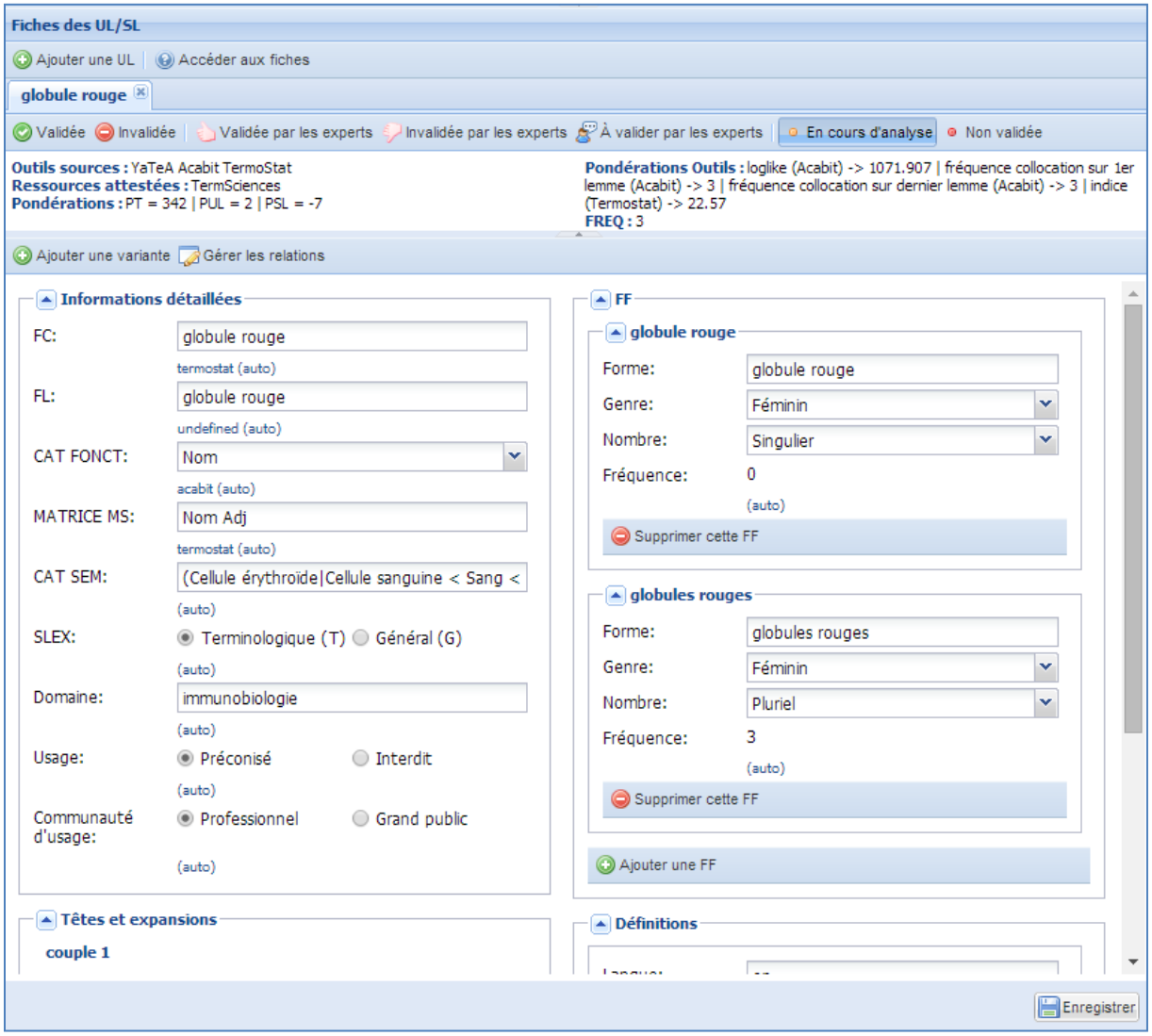


Figure 6-14. Fiche lexicale d'une UL

Afficher la fiche d'une UL/SL *Actions utilisateurs*

> dans l'espace *Liste des UL/ Liste des SL*, double-cliquer sur l'UL/SL à afficher ;

➔ la fiche lexicale de l'UL/ SL sélectionnée s'affiche dans l'espace *Fiche des UL/ SL*.

On peut ouvrir en même temps autant de fiches que nécessaire.

Ajouter une UL

Actions utilisateurs

> dans l'espace *Fiche des UL/SL*, cliquer sur *Ajouter une UL* ;

➔ une nouvelle fiche vierge s'affiche dans l'espace *Fiches des UL/SL* ;

> remplir la fiche > cliquer sur *Enregistrer* ;

➔ la nouvelle UL apparait dans la liste des UL (dans l'espace *Liste des UL*).

Ajouter une variante

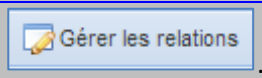
> dans l'espace *Fiche des UL/SL*, cliquer sur *Ajouter une variante* ;

➔ une nouvelle fiche vierge s'affiche dans l'espace *Fiches des UL/SL* ;

> remplir la fiche > cliquer sur *Enregistrer* ;

➔ la nouvelle variante apparait dans la fiche de relation de l'UL analysée (cf. Fonctionnalité *Gérer les relations*).

⚠ Une variante n'est pas considérée comme une UL autonome ; de ce fait, elle n'apparait pas dans la liste des UL, mais seulement dans la fiche de relation de l'UL dont elle est variante.

⚠ On peut accéder à la fiche de relation d'une UL (cf. Fonctionnalité *Gérer les relations*) directement à partir de sa fiche lexicale, en cliquant sur le bouton 

.

6.3.1 Contenu de la fiche lexicale d'une UL : modification et/ ou enrichissement manuel

La fiche lexicale d'une UL contient :

1. des informations non-modifiables, issues de l'analyse automatique du corpus (en haut de la fiche, cf. Figure 6-15) :
   - **Outils sources** : outils qui ont extrait l'UL analysée (parmi YaTeA, Acabit et TermoStat, cf. Choisir les outils) ;
   - **Ressources attestées** : ressources qui ont attesté l'UL, soit l'UL globale, soit sa Tête et/ou son Extension (parmi TermSciences et IATE ; cf. Choisir les ressources externes);
   - **Pondérations** : les trois indices de pondérations calculés par la Station pour l'UL analysée (cf. Paramétrer les pondérations);
   - **Pondérations outils** : pondérations spécifiques à chaque outil ayant extrait l'UL analysée (parmi Acabit et Termostat) ;
   - **FREQ** : fréquence de l'UL analysée (de toutes ses formes fléchies) en corpus, calculée par la Station.

Fiches des UL/SL
Ajouter une UL
Accéder aux fiches
globule rouge
Validée
Invalidée
Validée par les experts
Invalidée par les experts
À valider par les experts
En cours d'analyse
Non validée
Outils sources : YaTeA Acabit TermoStat
Ressources attestées : TermSciences
Pondérations : PT = 342 | PUL = 2 | PSL = -7
Pondérations Outils : loglike (Acabit) -> 1071.907 | fréquence collocation sur 1er lemme (Acabit) -> 3 | fréquence collocation sur dernier lemme (Acabit) -> 3 | indice (Termostat) -> 22.57
FREQ : 3

Figure 6-15. Informations non-modifiables dans la fiche lexicale d'une UL

2. des informations modifiables issues de l'analyse automatique du corpus ou ajoutées par l'analyste (cf. Analyse automatique), divisées en 5 encarts (cf. Figure 6-16) :
   - Informations détaillées ;
   - Têtes et expansions ;
   - FF (Formes Fléchies);
   - Définitions ;
   - SL issues de cette UL.

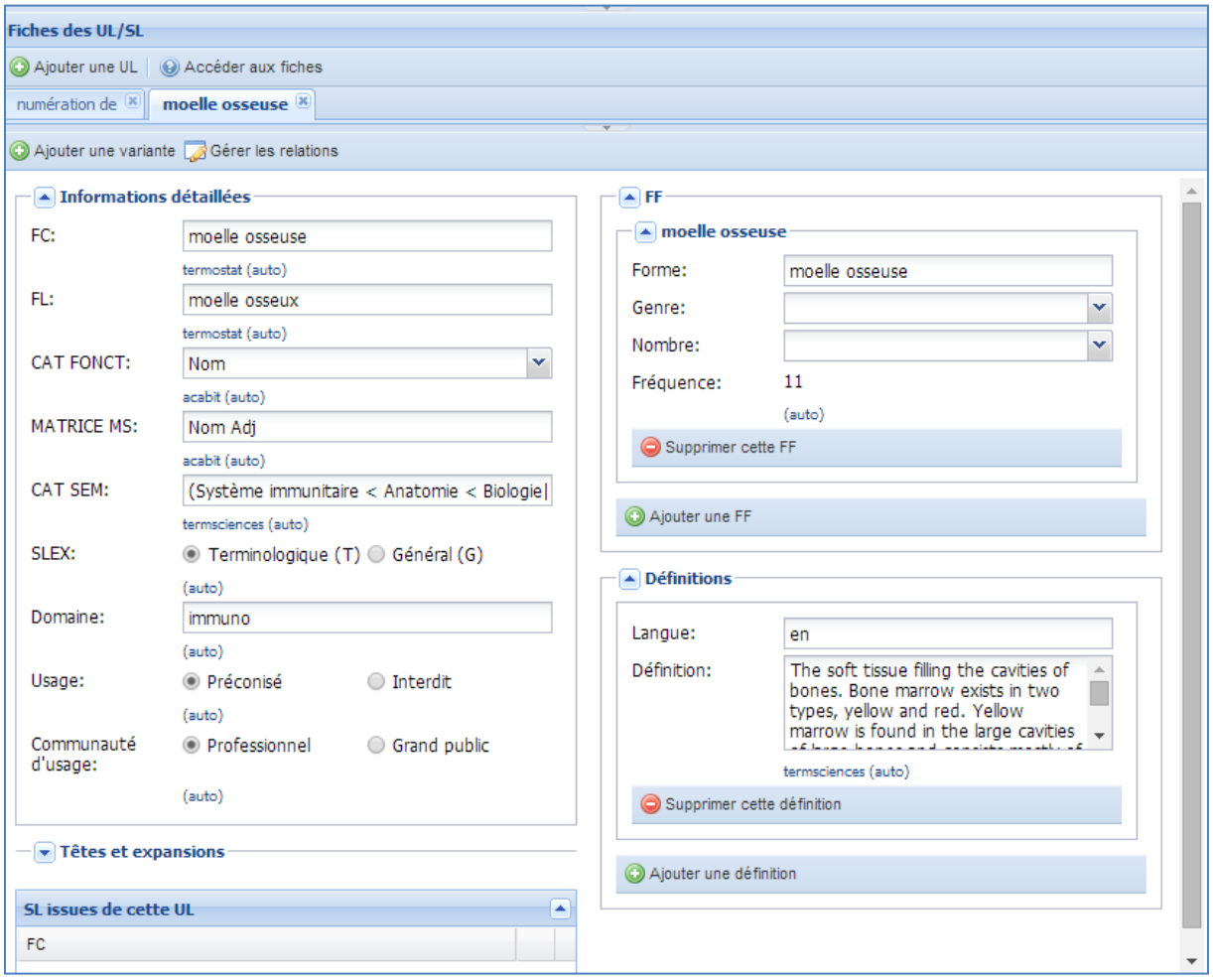


Figure 6-16. Les encarts dans la fiche lexicale d'une UL

Toutes les informations (sauf Fréquence) contenues dans ces encarts sont éditables, c'est-à-dire qu'elles peuvent être modifiées/ ajoutées/ supprimées soit librement, soit à partir d'une liste déroulante contenant des valeurs prédéfinies (cf. Figure 6-17). Par exemple, dans l'encart FF, le Genre doit être choisi parmi les 3 valeurs proposées : féminin, masculin, invariable, tandis que la Forme peut être modifiée librement.

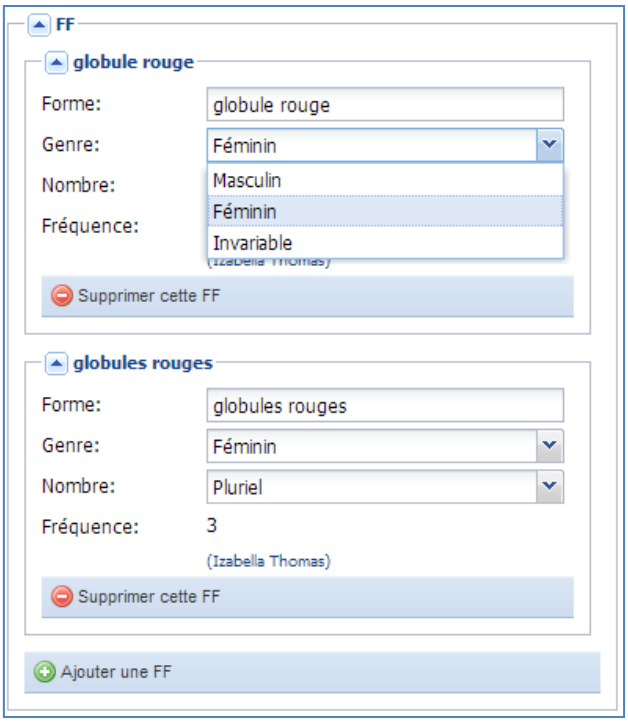


Figure 6-17. Informations modifiables dans la fiche lexicales d'une UL

Chaque action d'édition est tracée, c'est-à-dire assortie du nom de l'utilisateur/ outil l'ayant proposé, sous forme d'une petite étiquette bleue placée en dessous de chaque bloc d'informations (exemples : yatea (auto), termostat (auto), (Izabella Thomas)). L'étiquette (auto) renvoie le plus souvent au calcul effectué par la Station Sensunique.

Dans le cas de FF et Définitions, il est possible d'ajouter autant de nouvelles FF / Définitions que nécessaires ou de supprimer des FF/ Définitions existantes.

Pour être prise en compte, chaque action d'édition doit être enregistrée par un clic sur le bouton Enregistrer en bas de l'espace *Fiche des UL/SL*.

Les flèches permettent d'afficher / de cacher les blocs d'informations.

6.3.2 Contenu de la fiche lexicale d'une SL : modification et / ou enrichissement manuel

La fiche lexicale d'une SL contient (cf. Figure 6-18) :

1. des informations non-modifiables, issues de l'analyse automatique du corpus (en haut de la fiche) :
   - **Outils sources** : outils qui ont extrait la/les UL à la base de la SL (parmi YaTeA, Acabit et TermoStat, cf. Choisir les outils) ;
   - **Ressources attestées** : ressources qui ont attesté la/les UL à la base de la SL (parmi TermSciences et IATE ; cf. Choisir les ressources externes);
   - **FREQ** : fréquence de l'UL/ des UL à la base de la SL en corpus, calculée par la Station.
2. des informations modifiables issues du processus de dégradation (cf. Fonctionnalité Dégrader) ou ajoutées par l'analyste, divisées en 3 encarts :
   - Informations détaillées ;
   - Tête ;
   - UL à l'origine.

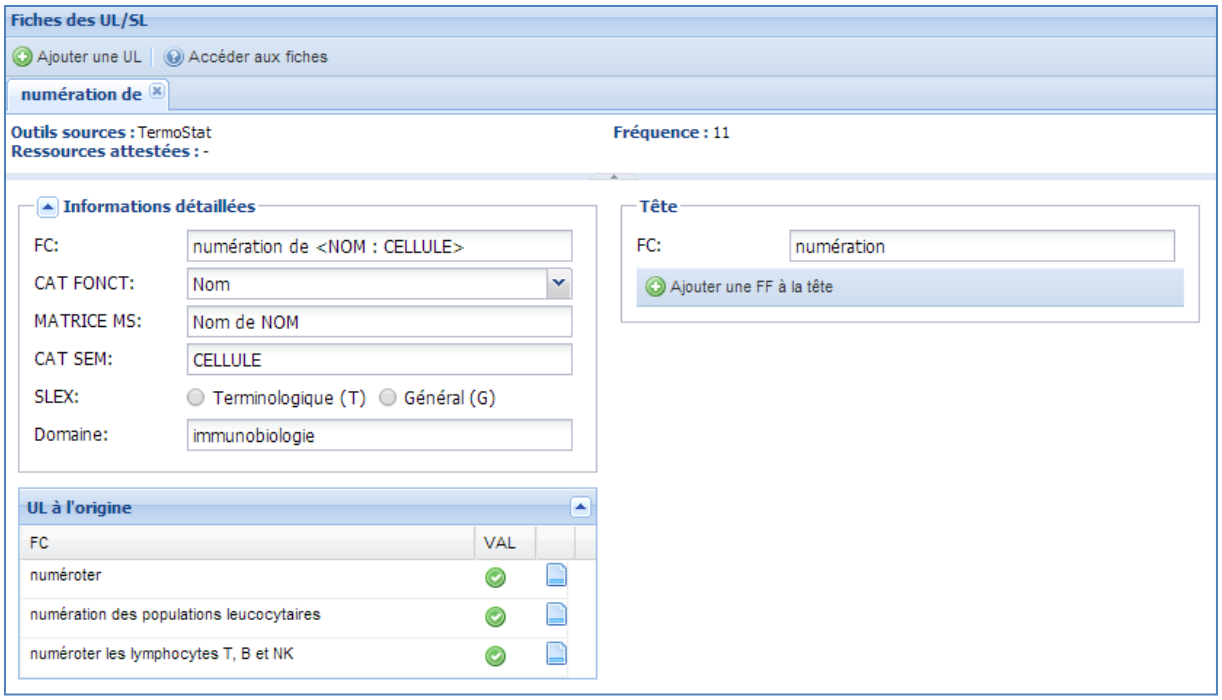


Figure 6-18. Fiche lexicale d'une SL

Comme pour la fiche lexicale d'une UL, toutes les informations contenues dans ces encarts sont éditables (sauf *UL à l'origine* de la SL), c'est-à-dire qu'elles peuvent être modifiées/ ajoutées/ supprimées soit librement, soit à partir d'une liste déroulante contenant des valeurs prédéfinies.

L'encart *UL à l'origine* permet aussi de visualiser les fiches lexicales des UL à l'origine de la SL.

Visualiser la fiche lexicale d'une UL à l'origine d'une SL *Actions utilisateurs*

> dans l'encart *UL à l'origine*, cliquer sur l'icône à côté de l'UL à visualiser ;

➔ la fiche lexicale de l'UL s'ouvre dans l'espace *Fiches des UL/SL*.

L'encart *Tête* permet d'ajouter une nouvelle FF de la Tête d'une SL, par exemple, pour la SL *numéroter < NOM : CELLULE>,* les FF du verbe *numéroter* (cf. Figure 6-19).

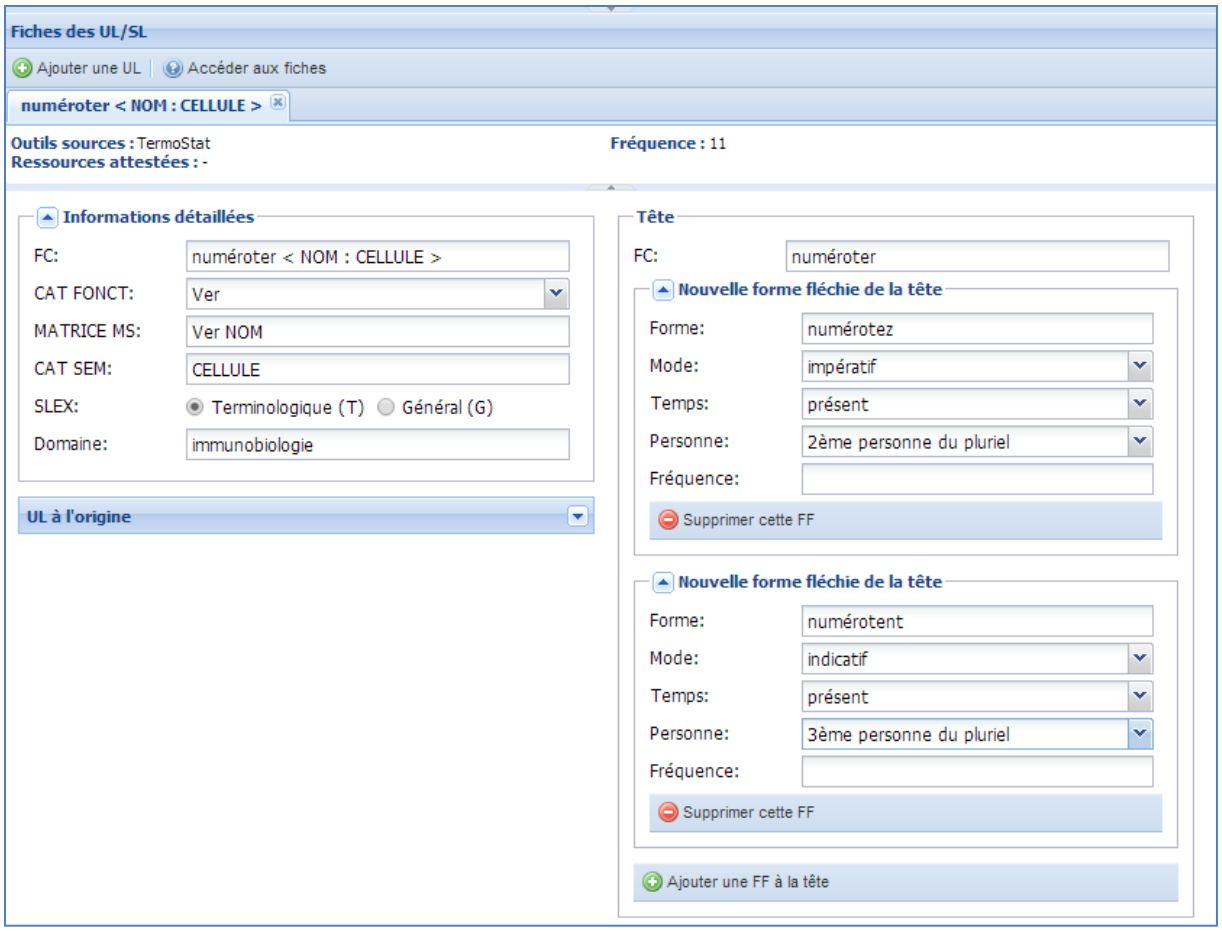


Figure 6-19. Ajout d'une FF de la Tête d'une SL

Pour être prise en compte, chaque action d'édition doit être enregistrée par un clic sur le bouton Enregistrer en bas de l'espace *Fiche des UL/SL*.

6.3.3 Processus de validation des UL/SL

L'objectif du processus de validation est d'accepter ou de refuser les propositions d'UL/ SL issues de l'analyse automatique ou manuelle. La validation peut se faire en 6 étapes, correspondant aux différentes phases du processus d'établissement d'un LLC ou d'un lexique terminologique. Il n'est pas obligatoire de passer par toutes les étapes de validation.

La validation d'une UL/ SL se fait à partir de sa fiche lexicale, en cliquant sur le bouton de validation affiché en haut de la fiche (cf. Figure 6-20).

Fiches des UL/SL
Ajouter une UL
Accéder aux fiches
globule rouge
Validée
Invalidée
Validée par les experts
Invalidée par les experts
À valider par les experts
En cours d'analyse
Non validée
Outils sources : YaTeA Acabit TermoStat
Ressources attestées : TermSciences
Pondérations : PT = 342 | PUL = 2 | PSL = -7
Pondérations Outils : loglike (Acabit) -> 1071.907 | fréquence collocation sur 1er lemme (Acabit) -> 3 | fréquence collocation sur dernier lemme (Acabit) -> 3 | indice (Termostat) -> 22.57
FREQ : 3

Figure 6-20. Validation d'une UL/ SL

Les valeurs possibles de validation d'une UL/ SL sont :

- *Non validée* : pas encore traitée, valeur par défaut (correspond à la valeur numérique de 0 lors de l'utilisation des filtres (cf. Filtres sur la liste des UL/SL);
- *En cours d'analyse* (valeur 1) ;
- *A valider par les experts* (valeur 2) ;
- *Validée par les experts* (valeur 3) ;
- *Validée* (définitivement)(valeur 4) ;
- *Invalidée* (et supprimée définitivement de la liste des UL/ SL) (valeur 5).

La valeur de validation activée à partir de la fiche lexicale, est de suite propagée sur la liste des UL/SL.

On peut aussi invalider les UL dans l'espace Liste des UL/ SL (cf. Fonctionnalité *Invalider*).

6.4 Espace 4 : Visualisation en contexte

Cet espace (cf. Figure 6-21) est utilisé dans deux objectifs :

- pour visualiser en contexte les UL sélectionnées dans la liste des UL (cf. Fonctionnalité *Projeter*) ;
- pour visualiser les résultats de la recherche effectuée avec le concordancier (cf. Concordancier évolué).

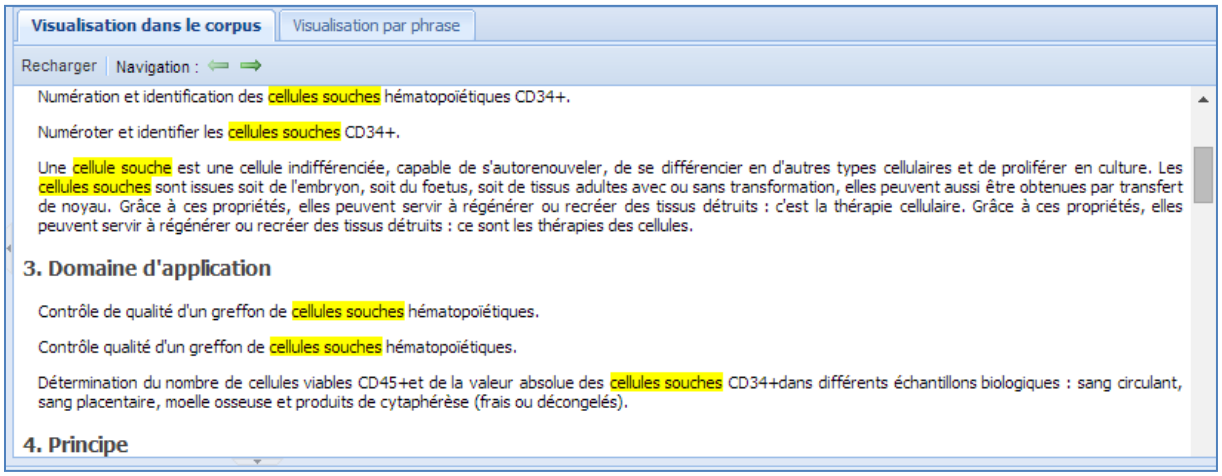

Visualisation dans le corpus
Visualisation par phrase
Recharger
Navigation :
Numération et identification des cellules souches hématopoïétiques CD34+.
Numéroter et identifier les cellules souches CD34+.
Une cellule souche est une cellule indifférenciée, capable de s'autorenouveler, de se différencier en d'autres types cellulaires et de proliférer en culture. Les cellules souches sont issues soit de l'embryon, soit du foetus, soit de tissus adultes avec ou sans transformation, elles peuvent aussi être obtenues par transfert de noyau. Grâce à ces propriétés, elles peuvent servir à régénérer ou recréer des tissus détruits : c'est la thérapie cellulaire. Grâce à ces propriétés, elles peuvent servir à régénérer ou recréer des tissus détruits : ce sont les thérapies des cellules.
3. Domaine d'application
Contrôle de qualité d'un greffon de cellules souches hématopoïétiques.
Contrôle qualité d'un greffon de cellules souches hématopoïétiques.
Détermination du nombre de cellules viables CD45+et de la valeur absolue des cellules souches CD34+dans différents échantillons biologiques : sang circulant, sang placentaire, moelle osseuse et produits de cytaphérèse (frais ou décongelés).
4. Principe

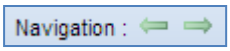
Figure 6-21. Visualisation en contexte

Les occurrences à visualiser sont mises en surbrillance.

6.4.1 Onglet Visualisation en corpus

Cet onglet permet de visualiser la sélection des UL / les résultats de la recherche, dans l'ensemble du Corpus d'Analyse.

Le bouton *Recharger* Recharger permet de rafraîchir le Corpus d'Analyse.

Le bouton *Navigation* 

permet de se déplacer d'une occurrence à l'autre, soit vers le haut ( ), soit vers le bas ( ) du corpus.

Un clic gauche sur une occurrence en surbrillance permet de passer d'une visualisation en corpus à la visualisation par phrase (de la même occurrence).

6.4.2 Onglet Visualisation par phrase

Cet onglet permet de visualiser la sélection des UL/ les résultats de la recherche dans les phrases isolées, extraites du Corpus d'Analyse. Chaque phrase affichée est identifiée par le nom du fichier dont elle est extraite et le numéro de la phrase (cf. Figure 6-22).

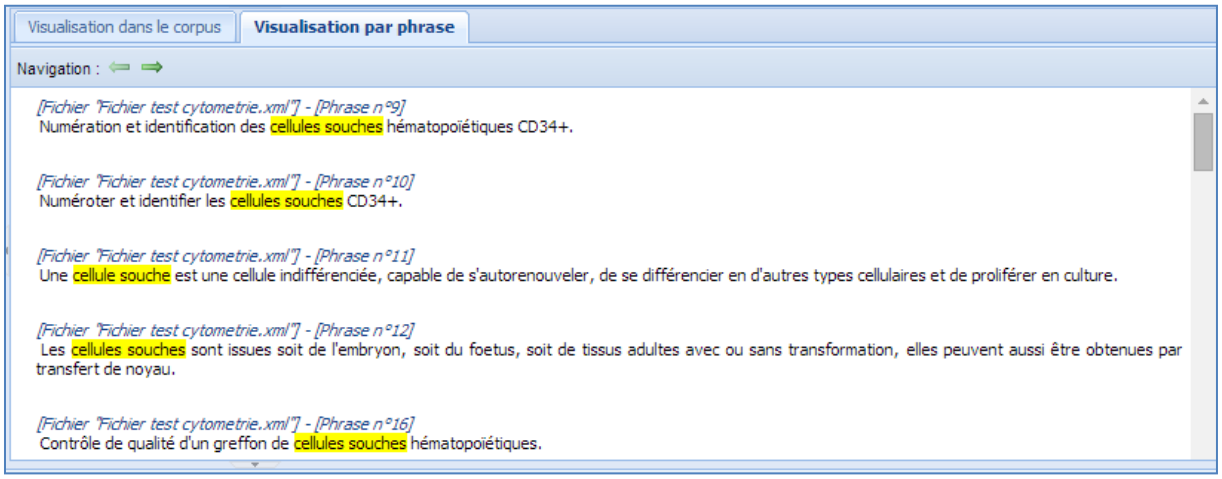

Visualisation dans le corpus
Visualisation par phrase
Navigation :
[Fichier 'Fichier test cytometrie.xml'] - [Phrase n°9]
Numération et identification des cellules souches hématopoïétiques CD34+.
[Fichier 'Fichier test cytometrie.xml'] - [Phrase n°10]
Numéroter et identifier les cellules souches CD34+.
[Fichier 'Fichier test cytometrie.xml'] - [Phrase n°11]
Une cellule souche est une cellule indifférenciée, capable de s'autorenouveler, de se différencier en d'autres types cellulaires et de proliférer en culture.
[Fichier 'Fichier test cytometrie.xml'] - [Phrase n°12]
Les cellules souches sont issues soit de l'embryon, soit du foetus, soit de tissus adultes avec ou sans transformation, elles peuvent aussi être obtenues par transfert de noyau.
[Fichier 'Fichier test cytometrie.xml'] - [Phrase n°16]
Contrôle de qualité d'un greffon de cellules souches hématopoïétiques.

Figure 6-22. Visualisation par phrase

Un clic gauche sur une occurrence en surbrillance permet de passer d'une visualisation par phrase à la visualisation en corpus (de la même occurrence).

6.5 Concordancier évolué

Le concordancier évolué offre différents types de recherche sur le corpus (cf. Figure 6-23) :

- **recherche simple** : sur une chaîne de caractères ;
- **recherche morphologique simple** : sur un (ou une suite de) lemme(s) permettant d'identifier toutes ses formes fléchies d'une UL ;
- **recherche morphologique complexe** : sur un (ou une suite de) radical(aux) permettant d'identifier les familles dérivationnelles ;
- **recherche morphosyntaxique** : sur une suite d'étiquettes morphosyntaxiques ;
- **recherche** dite **combinée** permettant de coupler les types de recherches précédents.

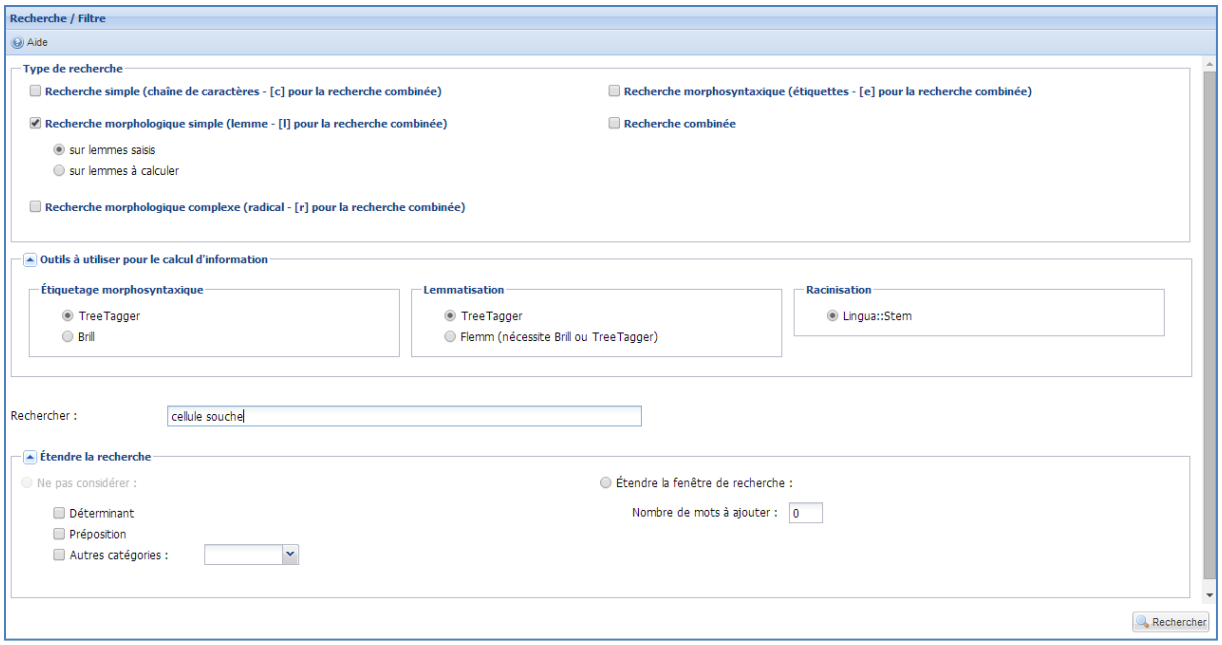


Figure 6-23. Concordancier évolué

Les différents types de recherche servent à comptabiliser les occurrences de motifs recherchés (cf. Onglet *Résultat de la recherche*) et de les visualiser en corpus (cf. Visualisation en contexte). Combiner des critères appartenant à différents niveaux d'analyse linguistique permet d'imposer des contraintes plus ou moins fortes sur les motifs recherchés, et ainsi cibler ou, au contraire, élargir le champ des résultats.

Les motifs de la recherche sont exprimés sous forme d'Expressions Régulières (selon http://fr2.php.net/manual/fr/book.pcre.php) adaptées à la Station Sensunique (cf. Expressions Régulières Sensunique).

Par exemple, la recherche combinée *[e]Nom [c]de [l]cellule* permet de cibler les groupes dont le premier élément est le Nom suivi de la préposition *de* et d'une des formes fléchies du mot *cellule* (ex. *nombre de cellules, greffon de cellules, analyse de cellules* etc.).

La recherche peut être faite sur les formes saisies manuellement (que ce soient des chaînes de caractères, des lemmes, des radicaux, des catégories morphosyntaxiques) ou sur les formes calculées par les outils, à partir des formes saisies.

Toute recherche peut être étendue (dans l'encart *Etendre la recherche*) :

- soit en ajoutant un certain nombre de mots aux résultats de la recherche ; par exemple, pour la recherche simple sur le motif *cellules souches* étendue d'un mot, on obtient les résultats suivants : *cellules souches*, *cellules souches hématopoïétiques*, *cellules souches CD34,* etc.

- soit en définissant des catégories de mots à ne pas considérer lors de la recherche d'un motif ; par exemple une recherche morphosyntaxique sur le motif *Nom Nom* avec l'omission d'une préposition, permet de retirer les occurrences de type *Nom Nom* (ex : *contrôle qualité*), mais aussi les occurrences de type *Nom Prep Nom* (ex : *contrôle de qualité*).

La recherche élargie permet de détecter divers phénomènes linguistiques, telles que la variation morphosyntaxique, la coordination ou plus généralement l'insertion d'éléments.

Les résultats de la recherche sont affichés sous l'onglet *Résultats de la recherche*.

Les occurrences sont visualisées dans l'espace Visualisation en contexte, en corpus (cf. Onglet *Visualisation en corpus*) et dans les phrases isolées (cf. Onglet *Visualisation par phrase*).

Effectuer une recherche à l'aide du concordancier *Actions utilisateurs*

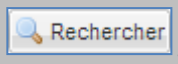
> cocher le type de recherche choisi dans l'encart *Type de recherche* ;

> si le Type de recherche choisi est *Recherche morphologique simple, Recherche morphologique complexe, Recherche combinée* :

> cocher la case correspondante au type des formes choisies : soit *formes saisies*, soit *formes à calculer ;*

> si la recherche s'effectue sur les formes à calculer, sélectionner les outils à utiliser pour le calcul dans l'encart *Outils à utiliser pour le calcul d'informations* ;

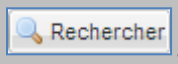
> saisir le motif de recherche dans l'encart *Rechercher ;*

> cliquer sur le bouton 

.

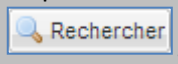
Etendre la recherche

Pour tout type de recherche :

> cocher *Etendre la fenêtre de recherche* dans l'encart *Etendre la recherche* > indiquer le nombre de mots à ajouter > cliquer sur le bouton 

.

Pour Recherche morphosyntaxique et Recherche combinée :

> cocher *Ne pas considérer* dans l'encart *Etendre la recherche* > cocher la catégorie morphosyntaxique des mots à ne pas considérer ou choisir la catégorie dans la liste déroulante > cliquer sur le bouton 

.

L'option *Ne pas considérer* est disponible seulement pour la Recherche morphosyntaxique et la Recherche combinée.

6.5.1 Expressions Régulières Sensunique

Pour tout type de recherche, sauf recherche combinée, les motifs de recherche sont exprimés sous forme d'Expressions Régulières (ER), selon http://fr2.php.net/manual/fr/book.pcre.php. Pour la recherche combinée, il faut utiliser les ER adaptées à la Station Sensunique.

Les recherches sont insensibles à la casse et à l'espace.

Recherche simple (sur une chaîne de caractères)

Exemples de motifs de recherche :

cellule ; Cellule ;

cellules souches; cellules Souches

Un message d'avertissement s'affiche lorsqu'aucune forme n'est trouvée (cf. Figure 6-24).

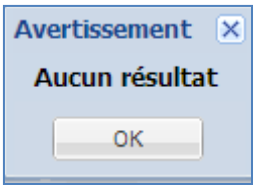


Figure 6-24. Message d'avertissement

Recherche morphologique simple (sur un (ou une suite de) lemme(s) permettant d'identifier toutes ses formes fléchies d'une UL)

Elle peut être effectuée :

- **sur les lemmes saisis** : saisir directement les lemmes

  Exemples de motifs de recherche :

  cellule ; analyse de cellule ; cellule tumoral ; utiliser

- **sur les lemmes à calculer** : saisir le motif à partir duquel les outils doivent calculer les lemmes

  Exemples de motifs de recherche :

  cellules ; analyses de cellules ; cellules tumorales ; utiliser

Un message sur les lemmes qui ont été effectivement calculés par les outils est retourné à l'utilisateur à chaque recherche, qu'elle soit aboutie (cf. Figure 6-25) ou non (cf. Figure 6-26).

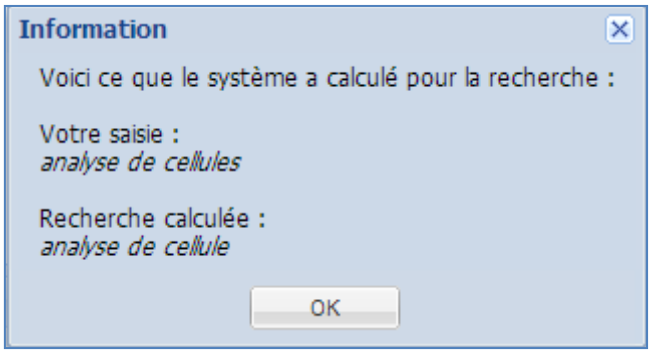


Figure 6-25. Message d'information sur les lemmes calculés par les outils

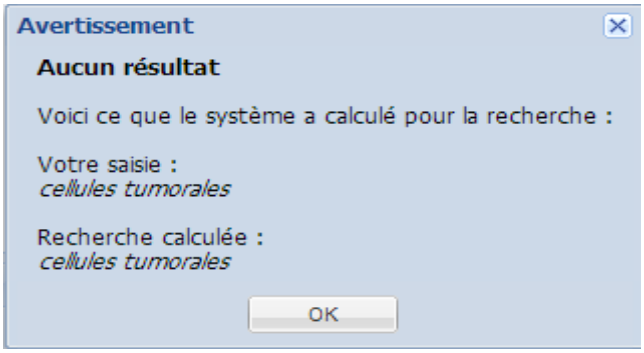


Figure 6-26. Message d'avertissement

Les lemmes calculés par les outils ne sont pas toujours linguistiquement corrects et ils peuvent varier en fonction des outils sélectionnés.

Recherche morphologique complexe (sur un (ou une suite de) radical(aux) permettant d'identifier les familles dérivationnelles)

Elle peut être effectuée :

- **sur les radicaux saisis** : saisir directement les radicaux

  Exemples de motifs de recherche :

  cellul ; analys de cellul ; cellul tumoral ; utilis

- **sur les radicaux à calculer** : saisir le motif à partir duquel les outils doivent calculer les radicaux

  Exemples de motifs de recherche :

  cellules ; analyses de cellules ; cellules tumorales ; utiliser

Un message sur les radicaux qui ont été effectivement calculés par les outils est retourné à l'utilisateur à chaque recherche, qu'elle soit aboutie (cf. Figure 6-27) ou non (cf. Figure 6-28).

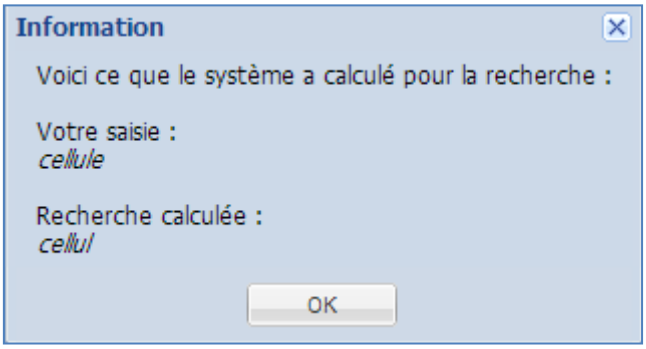


Figure 6-27. Message d'information sur les radicaux calculés par les outils

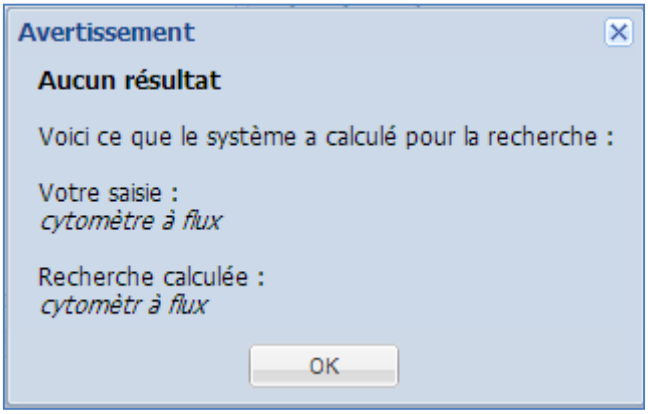


Figure 6-28. Message d'avertissement

Les radicaux calculés par les outils ne sont pas toujours linguistiquement corrects.

Recherche morphosyntaxique (sur une suite d'étiquettes morphosyntaxiques)

Pour effectuer une recherche morphosyntaxique, il faut utiliser le jeu d'étiquettes Sensunique (cf. Liste de catégories morphosyntaxiques Sensunique)

Exemples de motifs de recherche :

Nom ; Nom Prep Nom ; Nom Adj ; Ver Det Nom

Recherche combinée (permettant de coupler les types de recherches précédents)

Pour créer un motif pour une recherche combinée, chaque type de recherche utilisée doit être indiqué par une lettre minuscule entourée de crochets :

Expressions régulières Sensunique :

- **[c]** pour la recherche d'une chaîne de caractères
- **[l]** pour la recherche d'un lemme
- **[r]** pour la recherche d'un radical
- **[e]** pour la recherche d'une étiquette morphosyntaxique

Exemple :
[c]cellules [e]Adj? : permet de rechercher toutes les occurrences contenant la chaîne de caractères *cellules* éventuellement suivie d'un adjectif.

La **recherche combinée** peut être effectuée :

- **sur les formes saisies :**

  Exemples de motifs de recherche :

  [e]Nom [c]de [e]Nom : permet de rechercher toutes les occurrences contenant un Nom (recherche morphosyntaxique), suivi d'une chaine de caractères *de* (recherche simple), suivi d'un autre Nom (recherche morphosyntaxique), ex. *cytométrie de flux, nombre de cellules, produits de cytaphérèse,* etc.

  [r]utilis [l]de [e]Nom : permet de rechercher toutes les occurrences contenant le radical *utilis* (recherche morphologique complexe), suivi d'une forme dont le lemme est *de* (recherche morphologique simple), suivi d'un Nom (recherche morphosyntaxique), ex. *utilisation de monensine, utilisation d'anticorps,* etc.

- **sur les formes à calculer** :

  Exemples de motifs de recherche :

  [l]cellules [e]totales : permet de rechercher toutes les occurrences contenant un lemme calculé sur la chaîne *cellules*, i.e. *cellule* (recherche morphologique simple), suivi de la catégorie morphosyntaxique calculée sur la chaîne *totales*, i.e. Adj (recherche morphosyntaxique), ex. *cellules mononuclées, cellule productrice, cellules tumorales, cellules isolées,* etc.

  [r]utiliser [e]le [e]cytomètre : permet de rechercher toutes les occurrences contenant le radical calculé sur la chaîne *utiliser*, i.e. *utilis* (recherche morphologique complexe), suivi de la catégorie morphosyntaxique calculée sur la chaîne *le*, i.e. Det (recherche morphosyntaxique), suivi de la catégorie morphosyntaxique calculée sur la chaîne *cytomètre*, i.e. Nom (recherche morphosyntaxique), ex. *utiliser le sérum, utilisant les anticorps, utilisent le stem, etc.*

Un message sur les formes qui ont été effectivement calculées par les outils est retourné à l'utilisateur à chaque recherche, qu'elle soit aboutie (cf. Figure 6-29) ou non.

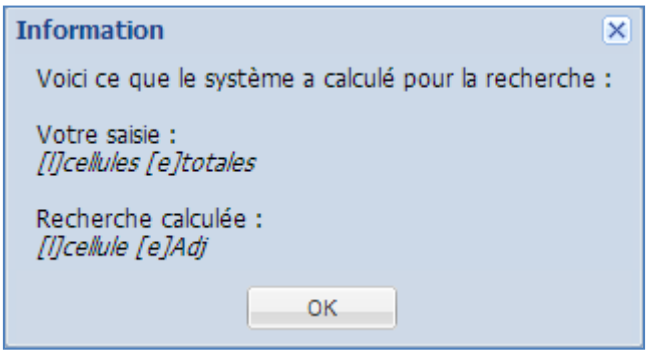


Figure 6-29. Message d'information sur les formes calculées par les outils

Comme pour toutes les autres recherches, les formes calculées par les outils ne sont pas toujours linguistiquement corrects.

7 Export d'un dictionnaire

Voir Fonctionnalité Exporter (UL) et Fonctionnalité Exporter (SL).

8 Quelques problèmes rencontrés lors de l'utilisation de la Station

L'analyse automatique est déclarée non-intègre.

Le plus souvent il s'agit d'erreurs dans les fichiers convertis, voir Prétraitement : préparation du corpus. Il faut vérifier, nettoyer, recharger le corpus et lancer une nouvelle analyse.

Le corpus ne s'affiche pas correctement dans l'espace de visualisation.

Dans la plupart de cas, il s'agit de problèmes d'intégrité d'analyse provoqués par des erreurs dans les fichiers convertis. Il faut vérifier, nettoyer, recharger le corpus et lancer une nouvelle analyse.

La Station ne répond pas, les actions ne sont pas exécutées.

Il faut vérifier si vous êtes toujours connecté à la Station. Après une période d'inactivité, la Station se déconnecte sans en prévenir l'utilisateur.

9 Annexes

9.1 Liste des abréviations

Forme abrégée	**Dénomination complète**
CA	Corpus d'Analyse
CAT FONCT	Catégorie Fonctionnelle
CAT MS	Catégorie Morphosyntaxique
CAT SEM	Catégorie Sémantique
CS	Corpus Support
CS	Corpus Contrastif
FC	Forme Canonique
FF	Forme Fléchie
FL	Forme Lemmatisée
FREQ	Fréquence
LC	Langue Contrôlée
LLC	Lexique d'une Langue Contrôlée
MATRICE MS	Matrice Morphosyntaxique
PSL	Poids de Structure Lexicale
PT	Poids Terminologique
PUL	Poids d'Unité Lexicale
SL	Structure Lexicale
SLEX	Statut Lexical
UL	Unité Lexicale
VAL	Validation
TAL	Traitement Automatique des Langues

9.2 Liste des catégories morphosyntaxiques Sensunique

Étiquette	Signification
Abr	Abréviations
Adj	Adjectifs
Adv	Adverbes
Conj	Cas de conjonctions de type indéterminé (coordination ou subordination)
Conjcoord	Conjonctions de coordination
Conjsub	Conjonctions de subordination (rattachés au verbe) + "que" ambigu
Det	Déterminants non contractés
Prepdet	Déterminants contractés ("du", "des", "au", "aux")
Inj	Interjections, onomatopées
Nom	Noms communs
Nomp	Noms propres (dont mots étrangers, potentiels noms propres)
Num	Numéraux cardinaux
Pna	Préfixes (exemple : anti) et autres particules non autonomes (exemple : -t-, quant (à))
Prep	Prépositions
Pro	Pronoms
Prorel	Pronoms relatifs (rattachés au nom)
Sym	Signes de ponctuation et symboles (exemple : "%")
Ver	Verbes aux autres modes
Vppe	Verbes au participe passé (dont participes passés adjectivaux, ou adjectifs participiaux)
Vppr	Verbes au participe présent
Zz	Mots non étiquetés ou ambigus ou problématiques

9.3 Listes des valeurs grammaticales

Mode :

'indicatif'

'conditionnel'

'impératif'

'subjonctif'

'infinitif'

'gérondif'

Temps par mode :

indicatif :

'présent'

'imparfait'

'plus-que-parfait'

'passé simple'

'passé composé'

'passé antérieur'

'futur simple'

'futur antérieur'

conditionnel :

'présent'

'passé 1ère forme

'passé 2ème forme'

impératif :

'présent'

'passé'

subjonctif :

'présent'

'imparfait'

'plus-que-parfait'

infinitif :

'présent'

'passé'

participe :

'présent'

'passé'

Personne :

'1ère personne du singulier'

'2ème personne du singulier'

'3ème personne du singulier'

'1ère personne du pluriel'

'2ème personne du pluriel'

'3ème personne du pluriel'

Genre :

'Masculin'

'Féminin'

'Invariable'

Nombre :

'Singulier'

'Pluriel'

'Invariable'

9.4 Exemple d'une ressource interne au format Sensunique

Source	id_UL	Forme canonique	Valeur de statut lexical			Domaine	Communauté d'usage		Langue	Catégorie sémantiqu		
Morphalou2.0		cheval	L		fr			Nom		chevaux	plur	mas
Morphalou2.0		cheval	L		fr			Nom		cheval	sing	mas
Morphalou2.0		porte	L		fr			Ver			sing	
Morphalou2.0		porte	L		fr			Ver			sing	
Morphalou2.0		porte	L		fr			Ver			sing	
Morphalou2.0		porte	L		fr			Ver			sing	
UMLF		monoclonal	L		fr			Adj		monoclonal		sin
UMLF		monoclonaux	L		fr			Adj		monoclonal		plu
Lefff-2.1		fronto-temporal	L			fr			Adj		fronto-temp	
Lefff-2.1		fronto-temporal	L			fr			Adj		fronto-temp	
Lefff-2.1		fronto-temporal	L			fr			Adj		fronto-temp	
Lefff-2.1		fronto-temporal	L			fr			Adj		fronto-temp	
RessourceLexicaleInventée		porte	L				fr			Nom		
RessourceLexicaleInventéeDérivation			utiliser		L			fr				Ver
RessourceLexicaleInventéeDérivation			utilisation		L			fr				Nom
Mon petit dictionnaire immunologique			Déficit immunitaire		T	Immunologie		Professionnels	fr	protéine		
Mon petit dictionnaire immunologique			CRP		T	Immunologie		Professionnels	fr	protéine		
RessourceTerminoInventéeEquitation			cheval		T			fr				
dela-fr-public-u8.dic		abasourdissement	L					fr			Nom	
dela-fr-public-u8.dic		abasourdissement	L					fr			Nom	
MeSH	D007153	Déficits immunitaires	T			fr						
MeSH	D007153	Déficits immunitaires	T			fr		allergologie et immunologie				
MeSH	D007153	Déficits immunitaires	T			fr		médecine interne				
MeSH	D007153	Déficits immunitaires	T			fr						
MeSH	D007153	Déficits immunitaires	T			fr						

Figure 9-1. Exemple d'une ressource interne au format Sensunique

Table des matières des figures

Figure 1-1. Architecture de la Station Sensunique 4
Figure 2-1. Oxgarage 8
Figure 2-2. Exemple d'un fichier converti en format TEI P5 contenant des caractères corrompus 9
Figure 3-1. Interface de connexion à la Station Sensunique 10
Figure 4-1. Gestion des utilisateurs 11
Figure 4-2. Interface de projet, Onglet *Utilisateurs* 13
Figure 4-3. Onglet Projets, vue générale 14
Figure 4-4. Onglet Outils et ressources pour la proposition d'UL 16
Figure 4-5. Valeurs de pondérations 19
Figure 4-6. Ressources externes 20
Figure 4-7. Ressources internes 21
Figure 4-8. Outils pour le concordancier 22
Figure 4-9. Onglet Corpus liés au projet 23
Figure 4-10. Onglet *Analyse* 24
Figure 6-1. Interface de travail 28
Figure 6-2. Concordancier évolué 28
Figure 6-3. Flèche sur la bordure de la fenêtre à afficher/cacher 29
Figure 6-4. Onglet Liste des UL 29
Figure 6-5. Fonctionnalités dans l'Onglet Liste des UL 30
Figure 6-6. Visualisation des UL 31
Figure 6-7. Création d'une SL 33
Figure 6-8. Export du dictionnaire d'UL 34
Figure 6-9. Fiche de relation d'une UL 36
Figure 6-10. Liste des SL 38
Figure 6-11.Export du dictionnaire d'une SL 39
Figure 6-12. Onglet Résultats 40
Figure 6-13. Onglet Statistiques de l'analyse 40
Figure 6-14. Fiche lexicale d'une UL 42
Figure 6-15. Informations non-modifiables dans la fiche lexicale d'une UL 43
Figure 6-16. Les encarts dans la fiche lexicale d'une UL 44
Figure 6-17. Informations modifiables dans la fiche lexicales d'une UL 45
Figure 6-18. Fiche lexicale d'une SL 46
Figure 6-19. Ajout d'une FF de la Tête d'une SL 47
Figure 6-20. Validation d'une UL/ SL 47
Figure 6-21. Visualisation en contexte 48
Figure 6-22. Visualisation par phrase 49
Figure 6-23. Concordancier évolué 50
Figure 6-24. Message d'avertissement 52
Figure 6-25. Message d'information sur les lemmes calculés par les outils 52
Figure 6-26. Message d'avertissement 52
Figure 6-27. Message d'information sur les radicaux calculés par les outils 53
Figure 6-28. Message d'avertissement 53
Figure 6-29. Message d'information sur les formes calculées par les outils 55
Figure 9-1. Exemple d'une ressource interne au format Sensunique 59

Tables des matières des tableaux

Tableau 4-1. Récapitulatif des droits selon le type d'utilisateur 13
Tableau 4-2. Paramètres de pondération du PT 17
Tableau 4-3. Paramètres de pondérations du PSL 18
Tableau 4-4. Paramètres de pondération du PUL 18